

MINICURSOS DA



VII ESCOLA REGIONAL DE
SISTEMAS DE INFORMAÇÃO

ORGANIZADORES

TIAGO CRUZ DE FRANÇA
ALEXANDRE LOUZADA
ALESSANDRO CERQUEIRA

**MINICURSOS DA VII ESCOLA REGIONAL DE SISTEMAS DE
INFORMAÇÃO (ERSI-RJ 2021)**

Porto Alegre
Sociedade Brasileira de Computação – SBC
2021

M665 Minicursos da VII Escola Regional de Sistemas de Informação (ERSI-RJ 2021) / Organizadores: Tiago Cruz de França, Alexandre Louzada, Alessandro Cerqueira. – Porto Alegre: Sociedade Brasileira de Computação, 2021.
1 e-book (309 p.) : il.

ISBN: 978-65-87003-72-6.

Parceria: UNIRIO, UFRRJ, FAETEC e UNIVERITAS.

1. Computação - Congressos. 2. Análise de sistemas. I. França, Tiago Cruz de. II. Louzada, Alexandre. III. Cerqueira, Alessandro. IV. Sociedade Brasileira de Computação.

CDD 004.068

Prefácio

A Escola Regional de Sistemas de Informação do estado do Rio de Janeiro (ERSI-RJ) é um evento que reúne profissionais, professores e estudantes interessados em aprender e discutir problemas, soluções e conceitos relacionados a Sistemas de Informação. Os minicursos são atividades de curta duração (quatro horas) que fazem parte da programação da ERSI-RJ. Eles abordam temas relacionados a Sistemas de Informação com objetivo de proporcionar ao público da escola um ambiente de aprendizagem e de discussão de tendências e desafios na área de Sistemas de Informação.

É o primeiro ano da escola após a pandemia do novo coronavírus que levou ao cancelamento do evento em 2020. Este ano (2021), foram selecionados seis minicursos entre as propostas recebidas. Além desses, mais duas propostas foram aceitas para serem publicadas como capítulos deste livro. A seleção foi realizada por um comitê de avaliadores. Todos professores e profissionais de Sistemas de Informação. Os critérios de seleção das propostas foram: relevância para o evento, expectativa de público, atualidade e conteúdo.

Os capítulos deste livro abordam: os conceitos e formas de detecção das *fake news* nos meios digitais de divulgação de notícias; uma linguagem de modelagem organizacional que trata o tema de estratégia e alinhamento de recursos humanos baseado em competências; o uso de linguagem simples como suporte à transparência e ao desenvolvimento das melhores práticas para transformação digital; o gerenciamento de dados praticado no âmbito do jornalismo de dados; a introdução a projetos de jogos com fins educacionais que inclui um canvas destinado ao Design de Jogos Educacionais Endógenos, o qual integra conteúdos e regras; os conceitos, definições, aplicações e pesquisa de Ética em Sistemas de Informação; os principais protocolos da camada de aplicação da Internet das Coisas (do inglês, *Internet of Things*) com uma ênfase prática nos protocolos CoAP e MQTT; e um exemplo de aplicação da linguagem R em bases de dados sobre COVID-19 e transporte público na cidade do Rio de Janeiro correlacionando os dados, reduzindo tendências e sazonalidades.

Acreditamos que este material será útil em aulas de Sistemas de Informação; em discussões sobre novas abordagens de pesquisa suportando trabalhos atuais e futuros; e para apoiar a prática profissional. Parabenzamos e agradecemos aos autores dos capítulos. Agradecemos também ao Comitê de Seleção de Propostas pela dedicação e eficiência; ao Comitê Editorial pelo empenho; e a CESI (Comissão Especial de Sistemas de Informação) da SBC (Sociedade Brasileira de Computação) pelo apoio para publicação deste livro. Por fim, agradecemos ao aluno da graduação em Sistemas de Informação da Universidade Federal Rural do Rio de Janeiro Matheus Nunes Ritton pela capa do livro. Este é a sua segunda colaboração com as capas do livros de minicurso da ERSI-RJ.

Tiago Cruz de França (UFRRJ)
Coordenador de Minicurso da ERSI-RJ 2021
Tadeu Moreira de Classe (UNIRIO)
Coordenador Geral da ERSI-RJ 2021

VII Escola Regional de Sistemas de Informação do Estado do Rio de Janeiro

01 a 03 de Setembro de 2021

Rio de Janeiro – RJ – Brasil

MINICURSOS

Promoção

Sociedade Brasileira de Computação – SBC

Coordenação da Escola Regional de Sistemas de Informação do Estado do Rio de Janeiro 2019

Tadeu Moreira de Classe – UNIRIO

Henrique Prado de Sá Sousa – UNIRIO

Coordenação da Trilha Principal (Artigos)

Jobson Luiz Massollar da Silva – UNIRIO

Eduardo Kinder Almentero – UFRRJ

Coordenação do Workshop de Acompanhamento de Trabalhos Científicos (WATC)

Carlos Roberto De Oliveira Junior – IFRJ

Claudio Miceli de Farias – UFRJ

Marco André Abud Kappel – CEFET/RJ

Coordenação de Minicursos

Tiago Cruz de França – UFRRJ

Alexandre Louzada – FAETEC
Alessandro Cerqueira – FAETEC

Coordenação de Painéis, Palestras e Oficinas

Rafael Elias de Lima Escalfoni – CEFET-RJ
Fabrício Barros Gonçalves – IFFluminense

Comitê de Seleção de Propostas de Minicursos

Alana Morais – UNIESP
Alessandro Cerqueira – Univeritas-RJ
Alexandre Louzada – FAETEC
Bruna Diirr – UNIRIO
Claudio Miceli de Farias – UFRJ
Danilo S. Carvalho – UFRJ
Eduardo Goncalves – ENCE/IBGE
Emanuele Jorge – IFRJ
Lívia Ruback – UFRRJ
Marcos Arrais – PUC-MG
Nilton Rizzo – UFRRJ
Paulo Alceu Rezende – UFJF
Rafael Bernardo Teixeira – UFRRJ
Rafael Costa – IFRJ
Rafael Elias Escalfoni – CEFET-RJ
Raimundo Costa Macário – UFRRJ
Renata Araujo – Universidade Presbiteriana Mackenzie
Sergio Manuel Serra da Cruz – UFRRJ
Tadeu Classe – UNIRIO
Tiago Cruz de França – UFRRJ

Comitê Editorial

Tiago Cruz de França – UFRRJ
Tadeu Moreira de Classe – UNIRIO

Arte da capa de Matheus Nunes Ritton – UFRRJ

Sumário

Combate Automático às Fake News nos Meios Digitais.....	5
Paulo Márcio Souza Freire, Argus Antônio Barbosa Cavalcante, Ronaldo Ribeiro Goldschmidt	
Modelagem da Estratégia e Alinhamento de Recursos Humanos Baseado em Competências.	41
Henrique Prado de Sá Sousa, Eduardo Kinder Almentero	
Transparência e Transformação Digital: O Uso da Técnica da Linguagem Simples.	86
Claudia Cappelli, Vanessa Nunes e Rodrigo Oliveira	
Técnicas e práticas de jornalismo de dados para aquisição e gerenciamento de dados em MySQL aplicadas ao domínio da violência contra a mulher.	114
Luciana Sá Brito, Alayne Duarte Amorim, André Viana Tardelli, Angélica Fonseca da Silva Dias, Juliana Baptista dos Santos França, Adriana Santarosa Vivacqua	
ENDO-GDC: Projetando Jogos Educacionais.	154
Geraldo Xexéo e Bernardo Taucei	
Pensando-fazendo Sistemas de Informação com Ética. Da pesquisa à engenharia, e vice-versa.	204
Luiz Paulo Carvalho, Flávia Maria Santoro, Rosa Maria M. Costa, Jonice Oliveira	
Desvendando a Camada de Aplicação na Internet das Coisas: Teoria, Prática e Tendências.	250
Vagner E. Quincozes, Silvio E. Quincozes e Juliano F. Kazienko	
A linguagem R na análise de dados: Um estudo de caso dos transportes públicos do RJ durante a pandemia da Covid-19	285
Julia Amaro Gonçalves Fagundes, Matheus Henrique de Sousa Oliveira e Vladimir Fagundes	

Capítulo

1

Combate Automático às *Fake News* nos Meios Digitais

Paulo Márcio Souza Freire, Argus Antônio Barbosa Cavalcante, Ronaldo Ribeiro Goldschmidt

Abstract

Combating Fake News (i.e., false news intentionally spread) is not a recent problem. However, its complexity has increased mainly due to the growth of volume and speed of news dissemination provided by the digital media of news distribution (e.g.: social networks, online newspaper, etc). In this scenario, computational approaches are becoming essential devices to combat this type of news. Thus, this chapter presents a study about the main computational approaches to combat Fake News, besides some comments on related areas and recent research on this theme.

Resumo

O problema de combater Fake News (isto é, notícias falsas veiculadas de forma intencional) não é recente. Contudo, sua complexidade vem aumentando em função do crescimento do volume e da velocidade de divulgação de notícias proporcionado pelos meios digitais de divulgação de notícias (por exemplo, redes sociais, jornais on-line, etc.). Diante deste cenário, abordagens computacionais que possam auxiliar no combate deste tipo de notícia estão se tornando cada vez mais necessárias. Assim sendo, o presente capítulo apresenta um estudo sobre as principais abordagens computacionais de combate às Fake News, além de comentar sobre áreas e pesquisas atuais relacionadas a este tema.

1.1. Introdução

O consumo de notícias on-line vem aumentando a cada dia [Vosoughi et al. 2017]. Dentre as razões para este crescimento pode-se destacar o fácil acesso e o baixo custo proporcionado pelos meios digitais de divulgação de notícias (*MDDN*), compostos, basicamente, pelas mídias virtuais (por exemplo, jornais on-line), redes sociais e aplicativos de troca de

mensagens) [Souza Freire et al. 2021]. A divulgação de uma notícia em um meio digital está relacionada com dois eventos (isto é, publicação e propagação) [Shu et al. 2017]. A publicação corresponde ao evento no qual a notícia é disponibilizada pela primeira vez em um meio digital. Por outro lado, toda e qualquer reação ocorrida após a publicação (por exemplo, um comentário ou compartilhamento) corresponde a um evento de engajamento e, conseqüentemente, de propagação da notícia.

Apesar de seus benefícios, alguns *MDDN*, tais como as redes sociais, permitem que qualquer pessoa, independentemente de sua credibilidade, divulgue notícias com intenso poder de propagação [Shu et al. 2017, Wang et al. 2018a]. Tal permissividade amplificou a disseminação de *Fake News*, um tipo particular de notícia falsa cuja divulgação acontece de forma proposital [Freire and Goldschmidt 2019, Conroy et al. 2015, Zhang et al. 2018]. A proliferação de *Fake News*, geralmente, afeta não apenas a integridade jornalística, mas também perturba as áreas social, política, econômica, cultural, assim como da saúde e segurança [Mejova and Kalimeri 2020, Mustafaraj and Metaxas 2017, Wang 2017].

Como exemplo do poder de influência desse tipo de notícia, pode ser destacado que nos três meses finais das eleições presidenciais americanas, realizadas em 2016, as notícias falsas publicadas na rede social Facebook, que favoreceram qualquer um dos dois candidatos, foram compartilhadas 37 milhões de vezes [Farajtabar et al. 2017]. Ademais, casos relacionados às *Fake News* não se limitam aos EUA. Conforme divulgado pela *BBC News*¹, após notícias falsas terem, supostamente, levado a linchamentos em 2018 na Índia, o *WhatsApp* anunciou um limitador para a quantidade de encaminhamentos de mensagem. Infelizmente, o Brasil também apresenta acontecimentos vinculados às *Fake News*. Segundo matéria divulgada pelo jornal *on-line G1*², uma dona de casa, de 33 anos, morreu dois dias após ter sido espancada por dezenas de moradores do município de Guarujá, no litoral de São Paulo. De acordo com essa matéria, ela foi agredida a partir de uma notícia falsa, divulgada por uma rede social, que afirmava que a dona de casa sequestrava crianças para utilizá-las em rituais de magia negra.

Além disso, as *Fake News* também podem trazer riscos à Segurança Nacional, haja vista a possibilidade de que notícias intencionalmente falsas consigam interferir nas atividades tanto da Inteligência quanto da Contraineligência. Ressalte-se que tais atividades são consideradas fundamentais e indispensáveis à segurança dos Estados, da sociedade e das instituições nacionais³. Portanto, há um apelo urgente para desenvolver estratégias efetivas para combater o impacto desse tipo de notícia falsa, não só no âmbito civil como também no militar, especificamente no que tange à Defesa Cibernética.

Como um último e atual exemplo do poder de influência das *Fake News*, pode-se destacar o caso da pandemia de *COVID-19* (*Coronavirus Disease-19*), causada pelo patógeno *SARS-Cov-2* e que já matou milhares de pessoas no ano de 2020 ao redor do mundo⁴, onde inúmeras *Fake News* têm sido divulgadas em *MDDN* [Mejova and Kalimeri 2020]. Estas divulgações têm dificultado de forma significativa o esclarecimento da população

¹BBC News - <https://www.bbc.com>

²G1 - <https://g1.globo.com>

³Agência Brasileira de Inteligência. <http://www.abin.gov.br/atividadeinteligencia/inteligenciaecontrainteligencia/>

⁴www.who.int/emergencies/diseases/novel-coronavirus-2019

sobre a disseminação da pandemia e sobre as devidas medidas de enfrentamento da doença a serem adotadas a cada instante.

Devido aos impactos das *Fake News* disponibilizadas em MDDN, diferentes segmentos da sociedade têm pesquisado como combatê-las [Zhou et al. 2019, UNESCO 2019, Flintham et al. 2018, Wang et al. 2018a, Campan et al. 2017, Kshetri and Voas 2017]. Como consequência, algumas ações mitigadoras estão sendo potencializadas, onde é possível enfatizar a criação de legislação punitiva⁵, os serviços de checagem de fatos (por exemplo, AosFatos⁶), as iniciativas educacionais presentes na alfabetização midiática [UNESCO 2016] (ex: Jogos Educacionais Digitais - JED que buscam capacitar pessoas para avaliar criticamente a veracidade das notícias [Passos et al. 2020, Passos et al. 2021]) e o emprego de abordagens computacionais nos MDDN [Freire and Goldschmidt 2020].

Dentre as ações mitigadoras, o emprego de abordagens computacionais vem se destacando devido à sua maior velocidade de atuação [Ruchansky et al. 2017]. Uma ação rápida se faz necessária, pois o espalhamento das *Fake News* se apresenta como um problema não trivial, tanto pelo volume de publicações quanto pela velocidade das suas respectivas propagações [Shu et al. 2017].

As abordagens computacionais utilizadas no combate automático às *Fake News* nos MDDN, conforme proposto por [Freire and Goldschmidt 2020], podem possuir duas funcionalidades: *Deteção e Intervenção*. Simplificadamente, enquanto a deteção procura identificar se uma notícia é intencionalmente falsa, a intervenção busca mitigar os efeitos da divulgação (publicação/propagação) dessa notícia em um determinado meio digital de divulgação de notícias.

Baseado nesta necessidade computacional, o presente capítulo provê uma introdução ao referido combate através da seguinte estrutura: a Seção 1.2 apresenta diferentes definições para o termo *Fake News*, assim como aborda o comportamento disseminativo deste tipo de notícia nos MDDN e conceituações sobre *Verdade*. Um levantamento sobre os trabalhos relacionados é realizado na Seção 1.3. Por fim, na Seção 1.4, são abordados os problemas em aberto.

1.2. Fundamentos

Apesar da histórica existência de notícias fraudulentas em nossa sociedade, a utilização do termo *Fake News* é relativamente recente. Desta forma, se faz necessária a caracterização das diferentes definições para *Fake News*, assim como o seu comportamento disseminativo e algumas das conceituações sobre *Verdade*.

1.2.1. Caracterização de *Fake News*

Apesar da originalidade da expressão, as *Fake News* não surgiram com o uso dos MDDN. Haja vista que, mesmo com as mídias tradicionais, já existiam pessoas que, por diferentes razões, divulgavam notícias falsas de forma proposital [Golbeck et al. 2018]. Independente do surgimento, devido à contemporaneidade do termo *Fake News*, surgiram

⁵<https://www12.senado.leg.br/noticias/materias/2020/06/02/nova-versao-de-lei-contr-fake-news-tera-restricoes-a-contas-anonimas-e-mais-poder-a-denuncias-de-usuarios>

⁶<https://www.aosfatos.org/>

diferentes definições, as quais podem ser organizadas em dois grupos.

O primeiro grupo considera que o aspecto proposital é fundamental, pois define as *Fake News* como publicações intencionais e verificadamente falsas [Conroy et al. 2015, Reis et al. 2019, Zhou et al. 2019, Campan et al. 2017, Wang et al. 2018a, Shu et al. 2017, Zhou and Zafarani 2018, Flintham et al. 2018, Mustafaraj and Metaxas 2017]. Assim, para esse grupo, não basta a notícia ser falsa para ser caracterizada como *Fake News*, é também preciso ter sido divulgada intencionalmente. Para enfatizar a diferença entre uma notícia falsa e uma intencionalmente falsa, pode-se utilizar dois termos denominados *misinformation* e *disinformation* [Golbeck et al. 2018, Campan et al. 2017]. Enquanto *misinformation* corresponde às notícias falsas divulgadas pela falta da informação verdadeira, a *disinformation* diz respeito às notícias falsas divulgadas com algum propósito. Com base nessas correspondências, para o primeiro grupo, é possível caracterizar *Fake News* como sendo uma *disinformation* [Kshetri and Voas 2017]. Cabe ressaltar que, apesar de pertencente ao primeiro grupo, o trabalho [Zhou and Zafarani 2018] é ainda mais específico em sua definição, pois só considera *Fake News* quando a notícia intencionalmente falsa é divulgada por uma agência de notícias.

Ademais, ainda de acordo com esse primeiro grupo, existem outros campos de estudo que, apesar de não se enquadrarem na área de *Fake News*, apresentam uma relação com o combate às notícias intencionalmente falsas. Alguns desses campos se encontram descritos abaixo:

- Classificação de Rumores (*Rumor Classification*) - Rumor é uma informação em circulação cuja veracidade não foi verificada no momento da publicação. Um rumor pode ser classificado como verdadeiro, falso ou ainda não verificado [Ma et al. 2015, Shu et al. 2017, Liu and Xu 2016, Vosoughi et al. 2017]. Portanto, uma *Notícia* não verificada antes da publicação é um *Rumor*, que pode ser caracterizado como *Fake News* a partir do momento que seja identificado como falso e intencional. A tarefa mais relacionada com o combate às *Fake News* é a classificação da veracidade dos rumores;
- Descoberta da Verdade (*Truth Discovery*) - é a descoberta da verdade de fatos conflitantes entre diferentes fontes [Shu et al. 2017, Li et al. 2015]. Assim, uma mesma *Notícia* pode conter afirmações diferentes (isto é, distintas opiniões), onde as intencionalmente falsas podem ser caracterizadas como *Fake News*. Assim, o combate às *Fake News* pode se beneficiar da Descoberta da Verdade para determinar a veracidade das afirmações;
- Detecção de Iscas de Cliques (*Clickbait Detection*) – procura identificar, nas páginas *Web*, as chamadas iscas de cliques que, praticamente, forçam o usuário a selecionar a opção apresentada. Nesse caso, o corpo do texto (*bodytext*) dos artigos é, frequentemente, pobre em relação ao seu cabeçalho (*headline*). Essa discrepância pode ser encontrada não só em *Clickbait*, como também em *Fake News*. Sendo assim, o *Clickbait* pode ser usado como um indicador de *Fake News* [Shu et al. 2017];
- Detecção de Bots (*Bot Detection*) – procura identificar o envio automático de informações nas redes sociais por meio de robôs [Braz and Goldschmidt 2017]. Esses

envios podem potencializar tanto a publicação quanto a respectiva propagação da *Fake News* [Wang et al. 2018a, Nasim et al. 2018, Ferrara et al. 2016];

- Checagem de fatos (*Fact Checking*) - são *Websites* ou *Frameworks* responsáveis pela verificação, normalmente realizada com a ajuda de especialistas, da veracidade de fatos divulgados em MDDN [Ciampaglia et al. 2015, Vo and Lee 2018, Sethi 2017, Ruchansky et al. 2017]. Inclusive, existem abordagens voltadas para a seleção automática de notícias a serem enviadas para a referida checagem de fatos [Kim et al. 2018, Tschitschek et al. 2018]. A verificação da verdade dos fatos pode ser utilizada na tarefa de detecção de *FakeNews* [Cazalens et al. 2018], assim como na criação de *datasets* [da Silva et al. 2020];
- Sistemas de Reputação (*Reputation System*) - são sistemas que buscam determinar o nível de confiança em MDDN baseados na obtenção de graus de reputação dos usuários [Vavilis et al. 2014, Hendrikx et al. 2015, Seo J. 2013, Deng et al. 2014, Sherchan et al. 2013]. A determinação de graus de reputação pode ser utilizada na tarefa de identificação das *Fake News*.

O segundo grupo, entretanto, tem uma definição mais genérica. Para esse segmento, as *Fake News* são todas as notícias falsas, independente da sua natureza intencional [Sharma et al. 2019, Castelo et al. 2019, Ajao et al. 2019]. Inclusive, consideram-se como *Fake News* outros tipos de notícia, como, por exemplo, Rumor.

Este trabalho adota a definição do primeiro grupo. Consequentemente, considera *Fake News* como sendo uma notícia intencionalmente falsa. A principal razão da escolha é que uma notícia propositalmente divulgada tende a ser mais bem elaborada, podendo causar mais malefícios aos usuários.

1.2.2. Comportamento Disseminativo

A disseminação e, conseqüente, divulgação de uma notícia se inicia pela sua publicação e provável propagação (Efeito de Câmara de Eco) [Shu et al. 2017]. Dessa forma, é importante destacar o momento no qual uma notícia pode ser caracterizada como *Fake News*. Basicamente, uma notícia intencionalmente falsa pode surgir de três formas. A primeira é quando a *Fake News* é iniciada (publicada) em uma mídia virtual, podendo, posteriormente, ser republicada em uma rede social ou em um aplicativo de troca de mensagens. Na segunda forma, esse tipo de notícia é iniciada (publicada) diretamente na rede social ou em um aplicativo de troca de mensagens. Independente de ter surgido ou não na rede social ou em um aplicativo de troca de mensagens, a partir da sua chegada, essa notícia pode ser potencializada pela sua propagação. A terceira é quando uma notícia não *fake* é publicada, porém se torna *fake*, a partir do seu espalhamento, de acordo com as contribuições intencionalmente falsas feitas durante a sua propagação.

Independente do momento de criação, a recente proliferação de notícias falsas e mal-intencionadas nos MDDN, em especial nas redes sociais, tem sido uma fonte de preocupação generalizada. Essa apreensão se deve pelo seu poder de espalhamento e, conseqüente, influência na sociedade [Flintham et al. 2018]. As razões que potencializam a divulgação das *Fake News* nos MDDN podem ser divididas em quatro categorias. A

primeira tem relação com poder de influência ocasionado pelos fatores inerentes ao ser humano, dentre eles podemos destacar que as pessoas [Shu et al. 2017]:

- Preferem receber informações que confirmem as suas opiniões sem, necessariamente, verificarem a veracidade da notícia;
- Tendem a aceitar as informações não pela análise da verdade, mas pela relação de ganhos e perdas que a notícia pode trazer para elas;
- Tendem a avaliar as informações sem a busca da veracidade, pois acabam acompanhando a aceitação dos outros.

A segunda categoria é a carência de legislação punitiva, uma das alegações para tal fato é que as referidas leis poderiam cercear a liberdade de expressão. A terceira categoria está vinculada ao potencial ganho financeiro com a divulgação de determinadas notícias nos MDDN [Kshetri and Voas 2017]. Já a quarta categoria advém da facilidade de criação de contas nas redes sociais [Conroy et al. 2015]. Um aspecto importante inerente à essa facilidade é a criação de contas digitais maliciosas por meio de divulgadores de natureza humana e/ou computacional [Shu et al. 2017]. Esses divulgadores subdividem-se em:

- *Bot* - robôs responsáveis por divulgar *Fake News*;
- Humano - pessoas (*trolls*) intencionadas em disseminar *Fake News*;
- *Cyborg* - mecanismos híbridos (*Bot/Humano*) que divulgam *Fake News*.

Ainda se tratando da facilidade de divulgação de notícias intencionalmente falsas nas redes sociais, uma das formas mais simples de criar uma *Fake News* é se infiltrar em uma comunidade de pessoas engajadas em discutir um determinado assunto. Para tanto, segundo [Mustafaraj and Metaxas 2017], devem ser realizados os seguintes passos para divulgação de *Fake News*: Criar um domínio falso (*website*), criar contas anônimas, identificar comunidades e usuários interessados em um determinado assunto, contaminar esses usuários com a notícia falsa e, finalmente, incentivar a discussão para que a *Fake News* seja espalhada.

1.2.3. Conceituações sobre Verdade

Conforme introduzido neste trabalho, os dois grupos que definem *Fake News* convergem ao considerá-las como falsas. Diante disto, torna-se importante procurar conceituar o que é *Verdade*.

Essa conceituação sobre *Verdade* de uma afirmação é ampla, podendo-se destacar a existência de diferentes concepções filosóficas sobre a natureza do conhecimento verdadeiro. Inclusive, tal amplitude pode ser caracterizada pelas concepções ceticista e relativista. A primeira concepção, na sua forma clássica (ou ceticismo pirrônico) defende uma postura suspensiva, haja vista que a multiplicidade de explicações acerca de uma mesma afirmação constitui, por si só, razão suficiente para nada se afirmar de forma absoluta [Empirico 1997]. A segunda concepção, baseada no relativismo absoluto, aceita

diferentes verdades para uma mesma afirmação [Rodrigues 2013]. Inclusive, é possível ressaltar a refutação tanto do ceticismo quanto do relativismo, pois é difícil para alguém declarar-se ceticista ou relativista sem se colocar fora ou acima de tal declaração. Isso acontece porque, se uma pessoa declara que "*não há verdade absoluta*" ou que "*todas as verdades são relativas*", aparece a dúvida se tal afirmação é ou não, respectivamente, cética ou relativa.

Em uma perspectiva jornalística, uma notícia pode ter a verdade constatada tanto pela verificação dos fatos quanto dos argumentos a respeito do seu conteúdo [Sponholz 2009]. Partindo da premissa de que um argumento consiste na justificativa de uma opinião, a agregação das opiniões (argumentações) pode ser uma forma de avaliar uma notícia como verdadeira ou não.

1.3. Trabalhos Relacionados

Uma revisão bibliográfica foi realizada em busca de sumarizar as evidências existentes acerca do combate automático às *Fake News* nos MDDN, assim como identificar lacunas no estado-da-arte a fim de sugerir áreas a serem mais investigadas (isto é, problemas em aberto). O processo de revisão bibliográfica foi realizado a partir de pesquisas em algumas bases de dados científicas que apresentam publicações atualizadas e relevantes, tais como: *ACM Digital Library*⁷, *IEEE Xplore*⁸, *Scopus*⁹ e *Science Direct*¹⁰. Nessas pesquisas foram normalmente considerados somente os trabalhos com foco principal no combate às *Fake News* escritos em inglês e português, cujo *qualis*¹¹ seja superior ou igual a *B4* e publicados a partir do ano de 2015. O ano de 2015 foi escolhido como período de corte, pois, até onde foi possível observar, os trabalhos de anos anteriores ainda não demonstravam resultados robustos no combate às notícias intencionalmente falsas (por exemplo, experimentos realizados exclusivamente com *datasets* criados para detecção de rumores). Cabe ressaltar que também foram considerados alguns artigos referenciados pelos trabalhos analisados.

Um dos produtos resultantes dessa revisão bibliográfica foi a criação de uma visão geral dos trabalhos vinculados ao combate automático às *Fake News*. Essa visão começa pela apresentação de um modelo comparativo que viabiliza a distinção entre abordagens computacionais voltadas para o referido combate [Freire and Goldschmidt 2020]. Em seguida, os trabalhos, juntamente com os seus respectivos conjuntos de dados (*datasets*), são brevemente descritos e enquadrados no citado modelo.

1.3.1. Modelo Comparativo

O combate automático às *Fake News* nos MDDN possui uma variedade de aspectos que podem ser considerados. Com o objetivo de facilitar a comparação e a consequente classificação das referidas abordagens, tais aspectos são categorizados na Figura 1.1. As próximas subseções detalham cada um desses aspectos.

⁷<https://dl.acm.org/>

⁸<https://ieeexplore.ieee.org/Xplore/home.jsp>

⁹<https://www.elsevier.com/pt-br/solutions/scopus>

¹⁰<https://www.sciencedirect.com/>

¹¹<https://qualis.ic.ufmt.br/>

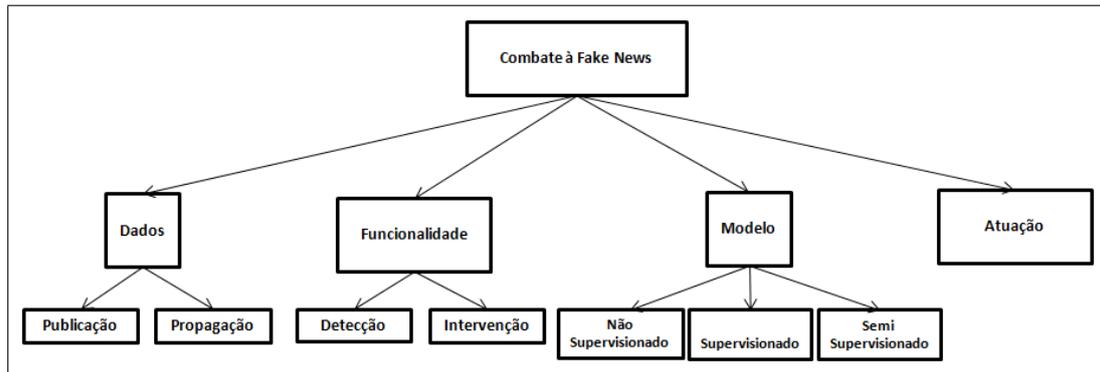


Figura 1.1: Aspectos considerados em Abordagens de Combate Automático às *Fake News*

1.3.1.1. Dados

Aspecto relacionado aos dados que podem ser utilizados pelas abordagens computacionais de combate às *Fake News*. Este aspecto subdivide-se em dados obtidos a partir da *Publicação* da notícia, como também aqueles associados com a sua *Propagação*.

Os dados de *Publicação* representam as informações inerentes ao surgimento da notícia no meio digital. Esses dados podem ser classificados em *Notícia*, *Usuário*, *Assunto e Temporalidade*. No que diz respeito à *Notícia*, a abordagem pode ser capaz de analisar dados oriundos da publicação a partir de diferentes tipos de *Mídia (Texto, Áudio e Imagem)*. Independente da *Mídia*, a análise do *Conteúdo* pode ser realizada de forma *Léxica, Sintática, Semântica e Legibilidade*. Com relação ao *Usuário* publicador, a abordagem pode identificar diferentes *Tipos*, tais como: humano, *bot* ou *cyborg*. Pode-se analisar também dados referentes ao *Perfil* do usuário na rede social, tais como: identificação e idade. Outro aspecto relevante está relacionado à *Reputação* do publicador, que pode estar vinculada à sua capacidade em identificar ou publicar *Fake News*. A abordagem pode também utilizar o *Assunto* abordado no momento da publicação. Assim, é possível tratar Especificidades, tais como: relacionamento entre assuntos, assuntos controversos ou análise de tópicos. Outro aspecto leva em consideração a *Relevância* do assunto publicado, uma vez que assuntos em voga motivam a criação de *Fake News*. A variação das características de uma notícia de acordo com o período de tempo da sua divulgação, torna a *Temporalidade* mais um relevante recurso para a identificação de *Fake News*.

Os dados de *Propagação* representam as informações obtidas após a publicação, consequentemente, aquelas inerentes às contribuições devido ao espalhamento da notícia na rede social (ex: curtida/like, comentário/reply ou compartilhar/retweet). Portanto, esses dados podem ser classificados em *Contribuição, Usuário, Assunto, Temporalidade e Rede*. No que diz respeito à *Contribuição, Usuário, Assunto e Temporalidade* a abordagem pode ser capaz de analisar os dados oriundos da *Propagação*, a partir dos mesmos aspectos anteriormente citados na *Publicação*. Ademais, as informações relacionadas à *Rede* criada, a partir da propagação da notícia, possibilitam não só a identificação de uma *Fake News* como uma possível atuação contra a mesma.

1.3.1.2. Funcionalidade

Além dos dados coletados, as abordagens automáticas de combate às *Fake News* podem, basicamente, possuir duas funcionalidades: *Deteção e Intervenção*.

A *Deteção* automática da *Fake News* pode ser, basicamente, um problema de classificação binária onde, dada uma rede social \mathcal{G} (caso particular de um meio digital de divulgação de notícia), uma notícia n é divulgada (publicada/propagada) por meio das suas respectivas postagens \mathcal{P}_n . Sendo que as divulgações pertencentes à \mathcal{P}_n são realizadas por um conjunto de usuários U_n em um instante t . Assim o referido classificador binário \mathcal{F} deve, aprendendo a partir dos dados, prever se n é uma *Fake News* ou não, como formalmente indicado na equação 1. Uma outra forma é a utilização de técnicas mais subjetivas que definam a probabilidade, peso ou pertinência de uma notícia n ser *fake*.

$$\mathcal{F}(n, \mathcal{P}_n, U_n, t) = \begin{cases} 1, & \text{se } n \text{ é uma } \textit{Fake News}; \\ 0, & \text{caso contrário.} \end{cases} \quad (1)$$

Independente da forma, para que uma notícia n possa ser detectada como *Fake News* é necessária a realização de duas subfuncionalidades: *Autenticidade e Intencionalidade* [Janze and Risius 2017, Vosoughi et al. 2017]. A *Autenticidade* analisa se a notícia é verdadeira ou falsa, enquanto que a *Intencionalidade* busca determinar a intenção dos divulgadores em ludibriar os receptores. Essa *Intencionalidade* pode ser mensurada como pontuação, peso ou score e obtida, por exemplo, por intermédio da análise de sentimentos que a notícia disponibiliza, pela associação entre usuários, assim como pelas características de perfil, tipo e reputação (credibilidade/confiança) dos divulgadores.

Já a *Intervenção* automática procura atacar as *Fake News*, nos MDDN, de forma proativa ou reativa [Shu et al. 2017, Farajtabar et al. 2017]. A intervenção reativa busca combater os efeitos da notícia a partir do momento da sua detecção como notícia propositalmente falsa. Por outro lado, a intervenção proativa tenta atuar antes mesmo da referida detecção, agindo então como uma forma de prevenção. Além disso, a tarefa de intervenção pode ser dividida em dois segmentos: *o Bloqueio e a Mitigação*. O *Bloqueio* atua de forma reativa. Na sua forma mais branda, o bloqueio interrompe a propagação da notícia e/ou a atuação do(s) usuário(s) divulgador(es). Uma outra forma mais incisiva seria remover a(s) notícia(s) e/ou o(s) usuário(s) divulgador(es). Já a *Mitigação* pode agir de forma reativa ou proativa buscando enfraquecer as consequências causadas pela *Fake News*. Na reatividade, a mitigação pode, por exemplo, imunizar os usuários provendo notícias verdadeiras [Farajtabar et al. 2017]. Uma forma de proatividade na mitigação é prover alertas, mesmo que a notícia ainda não tenha sido detectada como propositalmente falsa. Esses alertas podem estar relacionados com o nível de reputação da fonte (usuário) ou sobre o assunto estar relacionado com outras *Fake News* já identificadas.

Independente da funcionalidade da abordagem, a coleta dos dados inerentes à divulgação da notícia se faz necessária para subsidiar a detecção e a intervenção das notícias intencionalmente falsas. Assim, tanto a coleta de dados no meio digital quanto as tarefas de detecção e intervenção são fases iterativas, conforme ilustra a Figura 1.2. Cabe ressal-

tar que quanto mais cedo acontecer a detecção e a intervenção da *Fake News*, os impactos negativos dessa notícia tendem a ser menores.

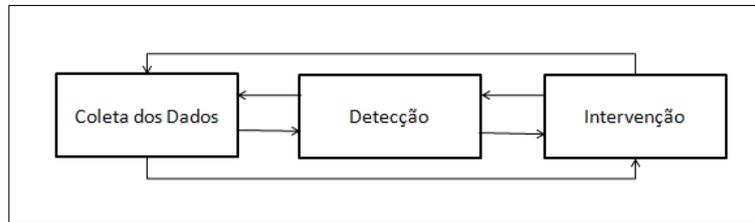


Figura 1.2: Fluxo do processo de combate às Fake News

1.3.1.3. Modelo

Quando a solução é por aprendizado de máquina, pode-se utilizar modelos computacionais para, a partir dos dados coletados, detectar as *Fake News*. Esses modelos são categorizados em *Não Supervisionado*, *Supervisionado* e *Semi-Supervisionado*.

No modelo *Não Supervisionado* são enquadradas as técnicas que normalmente levam mais tempo para realizar a identificação, porém, como não necessitam de rótulos, podem utilizar *datasets* mais simples [Shu et al. 2017].

Os modelos *Supervisionados* são lentos na fase do treinamento, entretanto, tendem a ser mais rápidos do que os não supervisionados no momento da sua utilização na identificação das *Fake News*. Devido à necessidade de treinamento, os modelos supervisionados precisam de *datasets* mais completos [Shu et al. 2017].

O modelo *Semi-Supervisionado* procura realizar a tarefa de identificação da *Fake News*, em MDDN, de uma forma mista que busque utilizar, tanto as técnicas supervisionadas quanto não supervisionadas. Essa abordagem pode utilizar *datasets* mais simples do que aqueles manipulados pelos modelos supervisionados, porém mais complexos do que os utilizados pelos não supervisionados [Shu et al. 2017].

1.3.1.4. Atuação

As abordagens computacionais que visam o combate às *Fake News*, independentemente dos dados coletados, funcionalidade e modelo utilizados, podem ter diferentes formas de atuação.

Uma das possibilidades de atuação está associada à localização física do combate dentro do meio digital. Uma abordagem *Centralizada* encontra-se, fisicamente, em um único ponto. Portanto, todas as tarefas relacionadas com a detecção/intervenção da *Fake News* são executadas em um mesmo local.

Por outro lado, uma abordagem *Descentralizada* encontra-se, fisicamente, espalhada. Assim, essa forma de atuação possibilita, inclusive, uma execução paralela e/ou distribuída [Wu and Liu 2018] no combate às *Fake News*.

1.3.2. Revisão dos Trabalhos

Nesta subseção são apresentados alguns trabalhos relacionados ao combate automático às *Fake News* nos MDDN. Para tal, foram realizadas buscas de acordo com a revisão bibliográfica já descrita, onde as principais fontes de consulta foram os artigos [Guo et al. 2020, Zhou and Zafarani 2018, Zhou and Zafarani 2019, Reis et al. 2019, Sharma et al. 2019] [Zhou et al. 2019, Shu et al. 2017, Conroy et al. 2015].

Para um melhor entendimento, os citados trabalhos são identificados e enquadrados no, já apresentado, Modelo Comparativo, conforme mostram as Tabelas 1.1, 1.2 e 1.3. Cabe ressaltar que, nessas três tabelas, as células não preenchidas indicam a não utilização do respectivo aspecto no trabalho correspondente.

Com base no referido enquadramento, uma constatação a ser destacada é a existência de um número significativo de abordagens computacionais que realizam a detecção de *Fake News*, nos MDDN, através da análise do texto existente na notícia. Uma possível razão para essa aparente preferência é que, inicialmente, as notícias intencionalmente falsas poderiam ser criadas sem uma forte preocupação de aparentar uma similaridade de conteúdo com as notícias não *fake*. Entretanto, atualmente, detecção através da análise do texto pode ser complexa, pois a cada dia os divulgadores intencionais de notícias falsas estão cada vez mais preocupados em manipular o conteúdo para que essas notícias pareçam não *fake* [Wu and Liu 2018, Liu and BrookWu 2018]. Sendo assim, as abordagens que não necessitam exclusivamente dos dados relativos ao conteúdo da notícia para a detecção de *Fake News* vêm se destacando. Dentre elas, as soluções baseadas na reputação dos usuários se apresentam como uma alternativa promissora, pois apresentam duas características interessantes. A primeira é a não necessidade de utilização do conteúdo da notícia, devido à, já citada, crescente similaridade entre as notícias *fake* e não *fake* que acaba por dificultar essa forma de detecção. A segunda é a não obrigatoriedade do uso dos dados relativos ao perfil do usuário na rede social, haja vista a dificuldade em se obter tais informações cada vez mais sigilosas [Shu et al. 2019b].

Além disso, os trabalhos são brevemente descritos, podendo seus detalhes serem consultados através das suas respectivas referências:

T1) *A Deep Transfer Learning Approach for Fake News Detection* [Saikh et al. 2020]: Esta pesquisa busca avaliar se o título de uma notícia está de acordo com o seu respectivo conteúdo utilizando *deep transfer learning*, onde o título é considerado como uma hipótese e a notícia é uma premissa. Assim, o objetivo é verificar se o corpo da notícia faz inferência ao seu respectivo título utilizando *stance detection*, tendo as seguintes respostas possíveis: *Agree, Disagree, Discuss e Unrelated*. Foram utilizados *Bi-LSTM e Bi-LSTM com max-pooling*. Os resultados foram comparados com a detecção humana, onde a acurácia máxima obtida foi de 0.90;

T2) *A Linguistic-Based Method that Combines Polarity, Emotion and Grammatical Characteristics to Detect Fake News in Portuguese* [de Souza et al. 2020]: O artigo apresenta o método (FNE) que, além da classificação gramatical e análise de sentimento baseada em polaridade, também usa a análise de emoções ao detectar notícias intencionalmente falsas escritas em português. Foram realizados experimentos com o FNE variando os classificadores *Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBo-*

ost (AB), Gradient Boost (GB) e Naive Bayes (NV). Os resultados foram comparados com os obtidos em [Moraes et al. 2019] (T12). A acurácia máxima obtida foi de 0.92;

T3) *A Topic-Agnostic Approach for Identifying Fake News Pages* [Castelo et al. 2019]: O trabalho propõe um *topic-agnostic* (TAG) classificador que usa dados linguísticos e *Web-Markup* (padrões de layout das páginas) para detectar *Fake News*. Assim, ao invés de usar o *bag of words*, o trabalho explora as *topic-agnostic*, incluindo características morfológicas, psicológicas e de legibilidade que são comuns em *Fake News*. O trabalho propõe que páginas com *Fake News* normalmente têm inclinação sensacionalista, assim como a ocorrência de termos, tais como: “*Just in*” e “*Read this*”. Foram utilizados 3 classificadores *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)* e *Random Forest (RF)*. Comparou o TAG com os resultados obtidos em [Pérez-Rosas et al. 2018] (T4), separando-os ano a ano (2013 até 2018). A acurácia máxima obtida foi de 0.86;

T4) *Automatic Detection of Fake News* [Pérez-Rosas et al. 2018]: Este trabalho cria uma ferramenta de detecção de *Fake News* por classificação com *Support Vector Machines (SVM)*, combinando informações léxicas, sintáticas, semânticas e de legibilidade. O presente trabalho compara os resultados com a detecção humana. A acurácia máxima obtida foi de 0.91;

T5) *Automatic Detection of Fake News on Social Media Platforms* [Janze and Risius 2017]: Este artigo implementa a detecção com os classificadores binários *Logistic Regression*, *Support Vector Machines (SVM)*, *Decision Tree*, *Random Forest* e *Extreme Gradient Boosting*. O referido trabalho compara os resultados entre os classificadores. A acurácia máxima obtida foi de 0.80;

T6) *Automatically Identifying Fake News in Popular Twitter Threads* [Buntain and Golbeck 2017]: O trabalho apresenta um método para detecção de *Fake News* no *Twitter* que acumula, ao longo do tempo, as características de rede, usuário e conteúdo para gerar uma regressão linear. Assim, a abordagem realiza a sua análise, levando em consideração os aspectos temporais relacionados à notícia a ser detectada. O artigo procura avaliar os resultados nos *datasets PHEME (Twitter para rumor)*, *CredBank (Twitter)* e *BuzzFeed News Fact-Checking Dataset (Facebook)* que precisaram ser alinhados com as mesmas características e rótulos. A acurácia máxima obtida foi de 0.70 nos experimentos realizados com o *dataset CredBank* para a detecção automática de *Fake News*;

T7) *BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection* [Zhang et al. 2020]: Compreende três módulos principais, sendo um extrator de recursos multimodais, um classificador de domínio e um detector de notícias falsas. Nesta proposta as características textuais são extraídas pelo modelo *BERT* e as características da imagem são obtidas pelo modelo *VGG-19*. Este método compara os seus resultados com outros trabalhos, como [Wang et al. 2018b] (T14). A acurácia máxima obtida foi de 0.85;

T8) *Beyond News Contents: The Role of Social Context for Fake News Detection* [Shu et al. 2019b]: Este artigo explora as correlações da postura da notícia, o bias e engajamento do usuário. Assim, é apresentado um Tri-Relacionamento (TriFN) onde tanto

informações partidárias quanto níveis de confiança do usuário podem ser utilizados para detecção de *Fake News*. Além disso, os usuários tendem a formar relacionamentos com pessoas afins que podem aumentar o espalhamento das *Fake News*. Essa abordagem compara os seus resultados com outros trabalhos, como [Rubin et al. 2015] (T32). A acurácia máxima obtida foi de 0.87;

T9) *CIMTDetect: A Community Infused Matrix-Tensor Coupled Factorization Based Method for Fake News Detection* [Gupta et al. 2018]: Através da modelagem de Câmara de Ecos, o trabalho representa uma notícia como um *3-mode tensor* $\langle \text{News}, \text{User}, \text{Community} \rangle$ e propõe um método baseado em *tensor factorization*. Além disso, apresenta uma extensão desse método com a junção de modelos que utilizam o conteúdo da notícia através de um *framework coupled matrix-tensor factorization*. Esse artigo usou o algoritmo de detecção da comunidade *Girvan-Newman* para identificar, na rede social, comunidades representativas de câmaras de eco. Os seus resultados são comparados com métodos que utilizam o classificador SVM, porém com diferentes formas de análise de conteúdo (ex. N-Gram). Os dois métodos propostos *CITDetect (community-infused tensor information)* e *CIMTDetect (community-infused tensor information + conteúdo da notícia)* utilizam o classificador SVM. O F1-score máximo obtido foi de 0.81;

T10) *Combining Neural, Statistical and External Features for Fake News Stance Identification* [Bhatt et al. 2018]: Neste estudo a ferramenta, desenvolvida para o primeiro desafio (FNC-1)¹², não tem o objetivo final de detectar se a notícia é *Fake News*. Nessa abordagem, as notícias são classificadas de acordo com a relação existente entre a manchete e o corpo do texto. Portanto os possíveis resultados da classificação são *Agree* - o texto do corpo concorda com a manchete, *Disagree* - o texto do corpo discorda da manchete, *Discuss* - o texto do corpo discute a mesma afirmação que o título, mas não toma uma posição ou *Unrelated* - o texto do corpo discute uma alegação que difere do título. A ferramenta combina as abordagens neural e estatística com recursos externos. Para isto, a solução implementa um modelo profundo recorrente (*Neural Embedding*), um modelo ponderado de características estatísticas (*n-gram bag-of-words*) e recursos externos criados à mão com a ajuda de uma heurística de engenharia de atributos. Por fim, usando uma rede neural profunda, todas as referidas abordagens são combinadas. Os resultados foram comparados com as demais ferramentas participantes do referido desafio (FNC-1). A acurácia máxima obtida foi de 0.89;

T11) *CSI: A Hybrid Deep Model for Fake News Detection* [Ruchansky et al. 2017]: O trabalho procura melhorar a acurácia na detecção de *Fake News* por meio de um modelo híbrido de rede neural profunda chamado CSI. Esse modelo utiliza três características: o texto da notícia, a resposta do usuário que recebeu a notícia e o usuário fonte da notícia. O CSI trabalha com o comportamento temporal dos usuários e da notícia. Esse modelo se divide em três partes: *Capture, Score e Integrate*. O primeiro módulo é baseado no texto e na resposta, por meio de uma rede neural recorrente (LSTM) para capturar um padrão temporal de atividades do usuário sobre a notícia e a representação *Doc2Vec*. O segundo usa uma rede neural para aprender as características da fonte, baseado nas interações dos usuários, gerando um score por meio de um grafo. Os dois módulos são integrados com o terceiro para caracterizar ou não a notícia como *Fake News*. O trabalho propõe a sua

¹²<http://www.fakenewschallenge.org/>

utilização em diferentes domínios, inclusive, em bancos de dados. Os resultados foram comparados com técnicas criadas para detecção de rumores. A acurácia máxima obtida foi de 0.95;

T12) *Data mining applied in fake news classification through textual patterns* [Moraes et al. 2019]: Este estudo realiza o levantamento de características textuais por meio do processamento do texto com um léxico de sentimentos. Assim, o método proposto calcula a polaridade total de cada texto que é adicionada ao conjunto de atributos (por exemplo, classes gramaticais) utilizados. Este método foi testado com os algoritmos de classificação *Support Vector Machines*, *Naive Bayes* e *Adaboost*. A acurácia máxima obtida foi de 0.93;

T13) *DistrustRank: Spotting False News Domains* [Woloszyn and Nejd1 2018]: Esta solução propõe uma estratégia de aprendizagem semi-supervisionada para separar automaticamente notícias falsas a partir de fontes não confiáveis de notícias. O trabalho utiliza como fonte *experts* de portais de checagem de fatos para classificar manualmente as notícias. A partir disto, é criado um grafo de pesos com os *ranks* de confiança sobre os *sites* e as arestas representam a similaridade dos mesmos. A pesquisa computa a centralidade, utilizando o *PageRank* em busca de uma similaridade entre os *sites* não confiáveis. O resultado da análise é a classificação em *Trust* ou *Distrust* para a fonte da notícia. O trabalho verificou que a semelhança entre os sites de notícias falsas é estatisticamente superior aos sites de notícias verdadeiras. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. O F1-score máximo obtido foi de 0.78;

T14) *EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection* [Wang et al. 2018b]: O artigo aponta que a maioria das abordagens existentes aprendem a detectar *Fake News* a partir de características específicas do evento, consequentemente, não podem ser transferidas para outros eventos ainda não aplicados. Assim, esse trabalho desenvolveu um *framework*, de ponta a ponta, denominado *EANN*, que pode derivar características invariantes de um evento para outro. Dessa forma, propõe uma detecção de *Fake News* para eventos recém-chegados. Isso consiste de três componentes principais: o extrator de características multimodais para texto e imagem (rede neural Convolutacional), o detector de *Fake News* (*fully connected layer com softmax*) e o discriminador de eventos (rede neural) que é o responsável por remover as características específicas do evento e manter as características compartilháveis entre os eventos para poder rotulá-los. Assim, o *framework* mede as características não similares entre diferentes eventos e remove-os para capturar as características invariantes entre eventos. Para avaliar seus resultados, realizou testes com técnicas de identificação de texto e imagem, porém utilizadas em trabalhos não ligados à detecção de *Fake News*. A acurácia máxima obtida foi de 0.82;

T15) *Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks* [Liu and BrookWu 2018]: O artigo propõe um modelo para detecção precoce de *Fake News* através da classificação dos caminhos de propagação da notícia. O referido trabalho modela o caminho de propagação de cada notícia como uma série temporal multivariada, na qual cada tupla é um vetor numérico que representa as características do usuário empenhado em espalhar a notícia. Para tal, é construído um classificador de série temporal que incorpora redes recorrente e

convolucional. Essas redes capturam as variações globais e locais das características do usuário, ao longo do caminho de propagação, para detectar *Fake News*. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.92;

T16) *Evaluating Machine Learning Algorithms for Fake News Detection*

[Gilda 2017]: Este artigo explora técnicas de linguagem natural para a detecção de *Fake News*. O trabalho aplicou *term frequency-inverse document frequency (TF-IDF)* de *bi-grams* e *probabilistic context free grammar (PCFG)* para um conjunto de 11.000 artigos em um *dataset* obtido pela *Signal Media*¹³ e uma lista de fontes da *OpenSources.com*¹⁴. Este *dataset* foi testado com os algoritmos de classificação *Support Vector Machines*, *Stochastic Gradient Descent*, *Gradient Boosting*, *Bounded Decision Trees* e *Random Forests*. A acurácia máxima obtida foi de 0.77, com modelos treinados apenas no conjunto de recursos do TF-IDF;

T17) *FActCheck: Keeping Activation of Fake News at Check* [Srivastava et al. 2018]:

Esta abordagem de intervenção sobre *Fake News* propõe uma melhoria na abordagem *competing cascades*, onde os *AFC* (*algoritmos polynomial time greedy*) e *RAFC* (*fast graph-pruning*) procuram escolher quais usuários têm maior poder de mitigação. Assim, os usuários com maior capacidade de influência na rede social realizam a mitigação através da divulgação de notícias alternativas (*Real News*). A máxima qualidade da redução obtida foi de 0.03;

T18) *Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning* [Souza Freire et al. 2021]:

Denominada *HCS* (*Hybrid Crowd Signals*), a abordagem proposta neste trabalho busca detectar se uma notícia *n*, divulgada em um meio digital, é *fake* ou não utilizando as reputações dos usuários (membros do *Crowd*) para ponderar as suas respectivas opiniões (*Signals*) sobre *n*. No entanto, a principal diferença entre a *HCS* e a abordagem baseada em *Crowd Signals* é a forma como as opiniões dos membros do *Crowd* são obtidas. Enquanto a abordagem baseada em *Crowd Signals* exige a opinião explícita dos usuários sobre os rótulos das notícias, a *HCS* considera obter as opiniões que se encontram *implícitas* nos padrões de comportamento dos usuários ao divulgar (publicar/propagar) essas notícias. Além disso, inspirada em meta-aprendizagem, a *HCS* permite ainda a formação de um *Crowd* híbrido, uma vez que os membros do *Crowd* podem ser tanto usuários divulgadores do meio digital quanto máquinas (modelos de classificação de notícias) disponibilizadas para o uso da *HCS*. O enquadramento da *HCS* no modelo comparativo, proposto neste capítulo, foi realizado somente com o método que utiliza as opiniões implícitas dos usuários (isto é, não utiliza as máquinas). A acurácia máxima obtida foi de 0.99;

T19) *Fake News Detection in Social Networks via Crowd Signals*

[Tschitschek et al. 2018]: A ferramenta desenvolvida trabalha na detecção e consequente intervenção de *Fake News*. Essa solução possui um algoritmo, chamado de *Detective* que usa inferência Bayesiana para detectar *Fake News* a partir de *Crowd Signals*. Esse *Crowd* é formado pela opinião dos usuários sobre a notícia, juntamente com a sua capacidade em opinar corretamente. O objetivo é detectar, de forma antecipada, a *Fake News* e bloqueá-

¹³<https://research.signal-ai.com/newsir16/signal-dataset.html>

¹⁴<http://www.opensources.co>

la. Os resultados foram comparados a partir de variações na própria abordagem, sendo as mesmas denominadas pelo artigo como Opt, Oracle, Fixed-CM e No-Learn. A máxima utilidade média obtida foi de 0.98;

T20) *Fake News Detection Using One-Class Classification*

[Faustini and Covões 2019]: O principal objetivo desse trabalho é eliminar a necessidade de contar com notícias rotuladas (como *fake* e não *fake*) na detecção de *Fake News*. Para que isso seja possível, seus autores utilizam o conceito de *One-Class Classification* (OCC). Nessa abordagem de aprendizado de máquina, o modelo de classificação é treinado utilizando apenas dados de uma classe, nesse caso, notícias rotuladas como *fake*, que costumam ser mais facilmente encontradas. Para avaliar essa abordagem no âmbito da detecção de *Fake News*, um algoritmo baseado em técnicas de redução de dimensionalidade foi estendido de modo a incorporar dados de publicação da notícia em formato de texto. O F1-score obtido a partir dos experimentos realizados variou entre 65% e 67%, onde os resultados foram comparados com abordagens similares;

T21) *Fake News Mitigation Via Point Process Based Intervention*

[Farajtabar et al. 2017]: Neste artigo, o enfoque está na intervenção de *Fake News*. A proposta é intervir, mitigando a notícia falsa, fornecendo recompensas na forma de notícias verdadeiras para quem recebeu a *Fake News*. A nível de influência da *Fake News* e a respectiva mitigação são quantificadas por contadores. O modelo utilizado foi baseado em *least-squares temporal difference learning* (LSTD). Um dos experimentos foi real, com a criação de cinco contas no *Twitter*. A melhora máxima de 0.2 foi obtida em relação a uma intervenção randômica;

T22) *FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization* [Shu et al. 2019a]: Apresenta o FakeNewsTracker, um sistema para detecção de notícias falsas. O FakeNewsTracker pode coletar, automaticamente, dados para notícias e contexto social. Esse trabalho propõe um *framework end to end* para realizar a coleta de dados, a detecção das *Fake News* e a visualização dos resultados. Essa pesquisa usa *autoencoders* para aprender o conteúdo de notícias e *RNN* para capturar o padrão temporal dos usuários de acordo com o seu engajamento com a notícia. O trabalho compara os seus resultados, internamente, a partir de variações do próprio *FakeNewsTracker*, onde são considerados somente o conteúdo da notícia ou o contexto social. Além disso, os resultados são comparados também com *Support Vector Machine*, *Logistic Regression* and *Naive Bayes*. A acurácia máxima obtida foi de 0.74;

T23) *IARNet: An Information Aggregating and Reasoning Network over Heterogeneous Graph for Fake News Detection* [Yu et al. 2020]: Sistema heterogêneo que considera a origem do post, os comentários e os respectivos usuários divulgadores, sendo estas informações armazenadas em um grafo. A detecção de *Fake News* utiliza *GloVe Embeddings*, *BERT representations*, *Bidirectional Gated recurrent units (BiGRU)* para o texto e *two-layer fully-connected MLP* para os dados do usuário. Esse trabalho compara os seus resultados com outros classificadores clássicos (por exemplo, *SVM* e *GRU*) a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.96;

T24) *Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection* [Wang 2017]: Além de propor um *dataset*, cria uma técnica de detecção de *Fake News* híbrida, usando redes neurais convolucionais (CNNs) para analisar, não somente textos,

mas também os dados do usuário. O artigo apresenta os resultados ao ser comparado com outros três detectores implementados com *Logistic Regression Classifier (LR)*, *Support Vector Machine Classifier (SVM)* e *bi-directional long short-term memory (Bi-LSTMs)*. A acurácia máxima obtida foi de 0.27;

T25) *Neural User Response Generator: Fake News Detection with Collective User Intelligence* [Qian et al. 2018]: O trabalho enfatiza a rápida propagação das *Fake News* nas redes sociais e, portanto, destaca a importância da sua detecção nos estágios iniciais, onde considera que apenas o texto da notícia está disponível. Tal afirmação se baseia no fato de que informações adicionais, como respostas dos usuários e padrões de propagação, podem ser obtidas somente após a notícia se espalhar. Contudo, como as respostas propagadas podem ajudar na tarefa de detecção, os autores propõem um *Two-Level Convolutional Neural Network with User Response Generator (TCNN-URG)* onde o TCNN captura a semântica do texto da notícia e o URG cria um modelo generativo de resposta dos usuários propagadores. O URG, a partir de respostas históricas, é treinado para aprender como os usuários respondem às notícias publicadas, gerando respostas de usuários para ajudar a TCNN na detecção da *Fake News*. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.89;

T26) *Ranking-based Method for News Stance Detection* [Zhang et al. 2018]: Mais uma pesquisa relacionada ao primeiro desafio (FNC-1). A solução do artigo é criada a partir de uma rede neural *Multi-Layer Perceptron*. Os resultados foram comparados com as demais ferramentas participantes do referido desafio. A acurácia máxima obtida foi de 0.86;

T27) *Real-time Detection of Content Polluters in Partially Observable Twitter Networks* [Nasim et al. 2018]: Esta pesquisa procura encontrar um tipo específico de *bots*, chamados de poluidores de conteúdo, para poder distinguir notícias verdadeiras de *Fake News*. Segundo o artigo, o estado da arte de detecção de *bots*, normalmente, necessita de um histórico completo da rede. Assim, o trabalho propõe uma abordagem baseada em informações parciais onde, ao invés de mapear um grafo com seguidores e seguidos, utiliza um grafo com a (dupla de Usuário) x (Evento). Essa dupla é obtida a partir do momento em que o par tenha *tweetado* no mesmo dia do evento. Dessa forma, os dados são clusterizados para que os usuários possam ser classificados como *bots* pela análise dos respectivos perfis e a frequência dos *tweets*. Os resultados do trabalho foram comparados com os obtidos por uma ferramenta citada pelo artigo, denominada de *Truthy*. A porcentagem de verdadeiros positivos obtida foi próxima de 0.65;

T28) *Sentiment Aware Fake News Detection on Online Social Networks* [Ajao et al. 2019]: O trabalho se aplica tanto a *Fake News* como Rumor. Assim, o artigo propõe a hipótese de que existe uma relação entre mensagens falsas ou rumores com os sentimentos dos textos. Foram utilizados dois modelos para extrair os escores de sentimento (positividade, negatividade ou neutralidade) do texto: *Latent Semantic Analysis (LSA)* e *Latent Dirichlet Allocation (LDA)*. O objetivo foi desenvolver um classificador que utilize os escores de sentimento. Assim, utilizando classificadores distintos, compara os resultados a partir da abordagem proposta com sentimentos. A acurácia máxima obtida foi de 0.89;

T29) *Sentiment-Based Multimodal Method to Detect Fake News* [Maia et al. 2021]: O artigo propõe o método (FNEXT) que considera o uso da análise de polaridade e da emoção extraídas tanto dos textos quanto das imagens existentes nas notícias para detecção de *Fake News* escritas em língua portuguesa. Foram realizados experimentos com o FNEXT variando os classificadores *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *AdaBoost (AB)*, *Gradient Boost (GB)* e *Naive Byes (NV)*. Os resultados foram comparados com os obtidos em [de Souza et al. 2020] (T2). A acurácia máxima obtida foi de 0.99;

T30) *Supervised Learning for Fake News Detection* [Reis et al. 2019]: Este artigo sobre *machine learning* explicável realiza uma pesquisa sistemática sobre *Fake News*, identificando os dados existentes, assim como propõe novos dados para detecção. Para implementar e avaliar esses recursos, o trabalho aplica os algoritmos de classificação *Extreme Gradient Boosting*, *Random Forest* e *SVM*. Ao aplicar os classificadores, os dados são subdivididos em características de conteúdo da notícia (por exemplo, texto e imagem), características da fonte (por exemplo, credibilidade e viés político), e características extraídas do ambiente (por exemplo, engajamento - likes e perfil do usuário). O trabalho propõe a utilização da combinação dessas características e comparou os seus resultados variando as características utilizadas por cada classificador de forma individual. A AUC máxima obtida foi de 0.88;

T31) *This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News* [Horne and Adali 2017]: Este trabalho usa um classificador SVM para detecção por meio da análise do texto, comparando os seus resultados entre detecção de *Fake News*, *Real News* e Sátira. Esse estudo determinou que as *Fake News* são mais próximas das Sátiras do que as notícias reais. A acurácia máxima obtida foi de 0.77;

T32) *Towards News Verification: Deception Detection Methods for News Discourse* [Rubin et al. 2015]: O trabalho propõe a ferramenta RST-SVM que analisa a notícia para extrair o estilo por meio da combinação do *Rhetorical Structure Theory (RST)* e *Vector Space Modeling (VSM)* para Clusterização. A detecção da notícia como enganosa ou real foi feita por meio de um classificador SVM. Os resultados obtidos foram comparados com a detecção humana. A acurácia máxima obtida foi de 0.56;

T33) *Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate* [Wu and Liu 2018]: Este trabalho busca a detecção de *Fake News*, pela modelagem da propagação da notícia através da mineração de grafos em Florestas. Segundo o artigo, classificar notícias pelo seu conteúdo é muito difícil, devido à similaridade entre as divulgações *fake* e não *fake*. Em contra partida, as *Fake News* tendem a ter as mesmas fontes e sequências. O trabalho propõe a ferramenta paralelizável chamada TraceMiner que utiliza *Recurrent Neural Networks (LSTM-RNNs)*, para classificar o caminho de propagação das mensagens no *Twitter*. O artigo comparou os seus resultados por meio de técnicas de análise de conteúdo criadas com SVM e XGBoost. O F1-score máximo obtido foi de 0.93;

T34) *Um Método Linguístico que combina Polaridade, Emoção e Aspectos Gramaticais para Detecção de Fake News em Inglês* [Testoni et al. 2021]: O artigo utiliza o já proposto método (FNE), porém propõe um protótipo adaptado que, além da classificação gramatical e análise de sentimento baseada em polaridade, também usa a análise de emoções ao detectar notícias intencionalmente falsas escritas em inglês. Foram realizados experimentos com o FNE variando os classificadores *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *AdaBoost (AB)*, *Gradient Boost (GB)* e *Naive Byes (NV)*. Os resultados foram comparados com os baselines sem a utilização da emoção. A acurácia máxima obtida foi de 0.99;

T35) *Weakly Supervised Learning for Fake News Detection on Twitter* [Helmstetter and Paulheim 2018]: Neste estudo, como existe uma dificuldade em conseguir um grande volume de dados para análise (*datasets*), os *tweets* são rotulados, automaticamente, durante a coleta, de acordo com a confiança na sua fonte. Assim é criado um *dataset*, denominado de *Large-scale Training Dataset*, onde cada tweet de uma fonte confiável é rotulado como uma notícia real, assim como, cada tweet de uma fonte não confiável é rotulado como uma *Fake News*. Espera-se que nesse *dataset* a classe de notícias reais contenha apenas uma quantidade negligenciável de ruído, pois fontes confiáveis raramente divulgam *Fake News*. Também é criado um segundo *dataset*, denominado de *Small-scale Evaluation Dataset*, possuindo *tweets* rotulados manualmente, como *fake* e não *fake*, a partir do site PolitiFact¹⁵. O objetivo principal do trabalho é treinar um classificador, a partir do primeiro *dataset*, para aplicá-lo no segundo *dataset*. Portanto, esse classificador, apesar de ter sido treinado em um *dataset* desenvolvido a partir da confiança, é utilizado para detectar *tweets fakes* e não *fakes* no segundo *dataset*. Portanto, o artigo considera que o classificador foi treinado e avaliado com alvos distintos. Para a referida detecção foram levadas em consideração as características do usuário (ex: engajamento e quantidade de seguidores), do tweet (ex: dia da semana, hora e texto), do tópico (assunto) e do sentimento. Como algoritmos de aprendizado foram usados o *Naive Bayes*, Árvores de Decisão, *SVM* e Redes Neurais. Além disso, foram usados dois *comitês*, o *Random Forest* e *XGBoost*. Os resultados foram comparados, utilizando diferentes combinações para os classificadores. O F1-score máximo obtido foi de 0.89;

T36) *XFake: Explainable Fake News Detector with Visualizations* [Yang et al. 2019]: O detector *XFake* é composto por 3 *frameworks*: *MIMIC*, *ATTN* e *PERT*. O *MIMIC* é construído para análise de atributos (ex. contexto da notícia e publicador) por meio de uma *deep neural network*. O *ATTN* é para análise semântica através de *pre-trained word embedding*, rede neural convolucional e *self-attention mechanism*. O *PERT* é para análise linguística utilizando um classificador *XGBoost*. A ferramenta, além de realizar as predições, também possui um módulo de interface para prover os usuários de explicações sobre as predições. O *XFake* é implementado em Python e deployed em *FLASK* com *front-end em HTML*. Para comparar seus resultados, o trabalho utilizou mão-de-obra humana realizada pela Amazon Mechanical Turk¹⁶. A acurácia máxima obtida foi de 0.67.

¹⁵<https://www.politifact.com/>

¹⁶<https://www.mturk.com/>

Tabela 1.1: Comparação entre abordagens - Dados de Publicação

Id	Dados Publicação							Temporalidade
	Notícia		Usuário			Assunto		
	Mídia (Texto, Áudio e Imagem)	Conteúdo (Léxica, Sintática, Semântica e Legibilidade)	Tipo (Humano, Bot e Cyborg)	Perfil	Reputação	Especificidades	Relevância	
T1	Texto	Léxica e Semântica						
T2	Texto	Léxica						
T3	Texto	Léxica e Semântica						
T4	Texto	Léxica, Sintática, Semântica e Legibilidade						
T5	Texto e Imagem	Léxica						
T6	Texto	Léxica e Semântica		X			X	
T7	Texto e Imagem	Léxica e Semântica						
T8	Texto	Léxica e Semântica		X	X			
T9	Texto	Léxica		X				
T10	Texto	Semântica						
T11	Texto	Léxica e Semântica		X	X		X	
T12	Texto	Léxica						
T13					X	Assuntos controversos		
T14	Texto e Imagem	Léxica						
T15				X			X	
T16	Texto	Léxica e Semântica						
T17								
T18					X			
T19					X		X	
T20	Texto	Léxica e Semântica						
T21								
T22	Texto	Léxica e Semântica		X			X	
T23	Texto	Léxica e Semântica		X				
T24	Texto	Léxica e Semântica		X		Relaciona Assuntos		
T25	Texto	Semântica						
T26	Texto	Semântica						
T27			Bot	X			X	
T28	Texto	Léxica e Semântica						
T29	Texto e Imagem	Léxica						
T30	Texto e Imagem	Léxica		X			X	
T31	Texto	Léxica e Sintática						
T32	Texto	Semântica						
T33					X			
T34	Texto	Léxica						
T35	Texto	Léxica e Semântica		X		Análise dos Tópicos	X	
T36	Texto	Léxica e Semântica		X		Análise dos Tópicos		

Tabela 1.2: Comparação entre abordagens - Dados de Propagação

Id	Dados Propagação							Temporalidade	Rede
	Contribuição		Usuário			Assunto			
	Mídia (Texto, Áudio e Imagem)	Conteúdo (Léxica, Sintática, Semântica e Legibilidade)	Tipo (Humano, Bot e Cyborg)	Perfil	Reputação	Especificidades	Relevância		
T1									
T2									
T3									
T4									
T5	Texto	Léxica						X	
T6	Texto	Léxica e Semântica		X			X	X	
T7									
T8				X	X			X	
T9				X				X	
T10									
T11	Texto	Léxica e Semântica		X	X		X	X	
T12									
T13									
T14									
T15				X			X	X	
T16									
T17								X	
T18					X				
T19					X		X	X	
T20									
T21				X			X	X	
T22				X					
T23	Texto	Léxica e Semântica		X					
T24									
T25	Texto	Semântica							
T26									
T27			Bot	X			X	X	
T28	Texto	Léxica e Semântica							
T29							X	X	
T30									
T31									
T32					X			X	
T33									
T34									
T35	Texto	Léxica e Semântica							
T36									

Tabela 1.3: Comparação entre abordagens - Modelo, Funcionalidade e Atuação

Id	Modelo			Funcionalidade				Atuação (Centralizada ou Descentralizada)
	Não Supervisionado	Semi Supervisionado	Supervisionado	Detecção		Intervenção		
				Autenticidade	Intencionalidade	Bloqueio (Reativa)	Mitigação (Proativa e Reativa)	
T1		X	X	X				Centralizada
T2			X	X	Análise de Sentimentos			Centralizada
T3			X	X				Centralizada
T4			X	X				Centralizada
T5			X	X				Centralizada
T6			X	X	Análise das características dos usuários			Centralizada
T7		X	X	X				Centralizada
T8		X		X	Pontuação de credibilidade para os usuários			Centralizada
T9			X	X				Centralizada
T10			X	X				Centralizada
T11			X	X	Score para os usuários			Centralizada
T12			X	X	Análise de Sentimentos			Centralizada
T13		X		X	Atribui pesos de confiança aos websites			Centralizada
T14			X	X				Centralizada
T15			X	X				Centralizada
T16			X	X				Centralizada
T17			X	X			Reativa	Centralizada
T18			X	X	Reputação do usuário			Centralizada
T19			X	X		X		Centralizada
T20			X	X				Centralizada
T21				X			Reativa	Centralizada
T22		X		X				Centralizada
T23			X	X				Centralizada
T24			X	X	Associação com o usuário			Centralizada
T25			X	X				Centralizada
T26			X	X				Centralizada
T27		X		X	identificação de bots			Centralizada
T28			X	X	Análise de Sentimentos			Centralizada
T29			X	X	Análise de Sentimentos			Centralizada
T30			X	X	Pontuação de credibilidade para as fontes			Centralizada
T31			X	X				Centralizada
T32		X		X				Centralizada
T33			X	X	Relação entre os usuários			Centralizada (pode ser paralelizada)
T34			X	X	Análise de Sentimentos			Centralizada
T35			X	X	Análise de Sentimentos			Centralizada
T36			X	X				Centralizada

Com o objetivo de apresentar alguns *datasets*, a Tabela 1.4 relaciona os trabalhos acima apresentados com os seus respectivos *datasets*. Em seguida, a Tabela 1.5 enquadra esses repositórios de acordo com os dados fornecidos por cada um deles. Esse enquadramento é realizado no apresentado Modelo Comparativo, restrito ao aspecto *Dados*. Cabe ressaltar que, na Tabela 1.5, as células não preenchidas indicam o não fornecimento do respectivo dado no *dataset* correspondente. Além disso, são utilizadas as seguintes formas de disponibilização dos dados:

- No Dataset: a informação está armazenada na própria base de dados;
- Link para o dado: a informação não está armazenada na base de dados, mas o *dataset* disponibiliza um link direto para o dado específico;
- Link para a notícia: Nesta caso, o *dataset* simplesmente disponibiliza o link para a notícia. Assim, se faz necessário o acesso à notícia original para a retirada das informações desejadas.

Com base no referido enquadramento dos *datasets*, apesar da relevância do problema de combate às *Fake News* nos MDDN, foi possível constatar que os repositórios que contêm dados reais ainda estão raramente disponíveis para download. Como consequência, algumas das pesquisas relacionadas ao combate às *Fake News* adaptou *datasets* originalmente criados para investigar outros problemas em redes sociais, como divulgação de

Rumor. Esses datasets adaptados, geralmente, não contêm informações importantes para a detecção de *Fake News*, como rótulos *fake* / não *fake*. Além disso, a maioria desses *datasets*, adaptados ou originalmente criados para detecção de *Fake News*, não descrevem a propagação das notícias nas redes sociais, como uma mesma notícia divulgada por vários usuários e várias notícias divulgadas por um mesmo usuário. Assim, não há um consenso sobre os *datasets* de referência para esse problema [Shu et al. 2017]. Outro fator complicador para a criação de *datasets* é a carência de informação, proveniente das redes sociais, para combate às *Fake News*. Tal carência acontece pois, muitas vezes essas informações são apagadas, impossibilitando a sua análise [Mustafaraj and Metaxas 2017].

Além disso, os referidos *datasets* são brevemente descritos, podendo seus detalhes serem consultados através das respectivas referências:

D1) *BS Detector* [Shu et al. 2017]: Este *dataset* é coletado de uma extensão de *browser* chamada *BS Detector* que foi desenvolvido para checagem da veracidade de notícias. Os rótulos existentes são *Fake news*, *Satire*, *Extreme bias*, *Conspiracy theory*, *Rumor mill*, *State news*, *Junk science*, *Hate group* e *Clickbait*;

D2) *BuzzFace* [Santia and Williams 2018]: Este repositório foi criado pela equipe do BuzzFeed. Ele contém 2.282 artigos rotulados como *Mostly true*, *Mixture of true and false*, *Mostly false* e *No factual content*;

D3) *BuzzFeedNews (2016-10-facebookfact-check modificado)* [Janze and Risius 2017]: Conjunto de dados criado a partir do *BuzzFeedNews (2016-10-facebookfact-check)* (D4), sendo os artigos rotulados como *Fake* e *Non-Fake*;

D4) *BuzzFeedNews (2016-10-facebookfact-check)* [Shu et al. 2017]: Este *dataset* compreende as notícias, do *Facebook*, oriundas de nove agências para a eleição presidencial americana de 2016. Os eventos e artigos ligados foram checados por jornalistas do *BuzzFeed*. Ele contém 1.627 artigos rotulados como *Mostly true*, *Mixture of true and false*, *Mostly false* e *No factual content*;

D5) *Celebrity* [Pérez-Rosas et al. 2018]: Este *dataset* fornece os dados da notícia para análise de texto. As notícias verdadeiras e falsas foram retiradas da *Web*, sendo relacionadas com assuntos de celebridades;

D6) *CredBank* [Shu et al. 2017]: Conjunto de dados criado a partir do cruzamento de várias fontes, com aproximadamente 60 milhões de *tweets*, que cobrem 96 dias, iniciados em outubro de 2015. Todos os *tweets* são relacionados com mais de 1.000 eventos de notícias. Cada evento foi avaliado por 30 anotadores da *Amazon Mechanical Turk*. Os rótulos existentes são [-2] *Certainly inaccurate*, [-1] *Probably inaccurate*, [0] *Uncertain (doubtful)*, [+1] *Probably accurate* e [+2] *Certainly accurate*;

D7) *DataSet Emergent* [Zhang et al. 2018][Bhatt et al. 2018]: Neste repositório, as notícias são rotuladas como *Agree* (o texto do corpo concorda com a manchete), *Disagree* (o texto do corpo discorda da manchete), *Discuss* (o texto do corpo discute a mesma afirmação que o título, mas não toma uma posição) e *Unrelated* (o texto do corpo discute uma alegação que difere do título). Esta base faz parte do primeiro desafio (FNC-1) e foi criada a partir do *dataset* para detecção de rumor chamado *Emergent*;

D8) *DistrustRank Datasets* [Woloszyn and Nejd1 2018]: Foram desenvolvidos dois *datasets*. O primeiro, gerado com sites confiáveis, por meio do *SimilarWeb*¹⁷, tem 502 domínios e 396.422 *URLs* de notícias. O segundo, obtido com sites não confiáveis, através do *Wikipedia's list of prominent Fake News*¹⁸, possui 47 domínios e 37.320 *URLs* de notícias;

D9) *Facebook para Detective* [Tschatschek et al. 2018]: Repositório que considera os círculos sociais do *Facebook*, consistindo de 4.039 usuários (nós) e 88.234 arestas;

D10) *Factck.Br* [Moreno and Bressan 2019]: *Dataset* composto por textos de 1.313 notícias no idioma português que foram verificadas individualmente pelas agências de checagem de fatos brasileiras 'Lupa', 'Aos Fatos' e 'Truco'. Possui como principal característica o fato de ser atualizável, ou seja, há uma API que possibilita a atualização do *Dataset* com novas notícias analisadas por agências de checagem. As notícias são rotuladas em *Ainda é cedo para dizer, de olho, discutível, distorcido, exagerado, impossível provar, impreciso, insustentável, verdadeiro, falso e outros*;

D11) *FakeBr* [Monteiro et al. 2018]: Um dos primeiros *datasets* disponibilizados para viabilizar a construção de modelos de detecção de *Fake News* no idioma português. Possui 7.200 notícias, das quais 3.600 são rotuladas *verdadeiras*, obtidas em versões digitais de mídias virtuais tradicionais. As outras 3.600, obtidas em sites considerados propagadores de notícias falsas, são rotuladas como *fake*;

D12) *FakePedia* [Charles and Oliveira 2018]: *Dataset* composto por textos de notícias no idioma português que foram verificadas através de *sites* brasileiros: 'e-farsas' e 'boatos.org'. As notícias são rotuladas em *fake news ou true news*;

D13) *Fake News vs Satire* [Golbeck et al. 2018]: *Dataset* para diferenciar *Fake News* e Sátiras onde as notícias são codificadas manualmente. A base, oriunda de diversas fontes, é composta por 283 relatos rotulados como *Fake News* e 203 como *Satirical*. Estes relatos são compostos pelo título, texto e um link para cada artigo;

D14) *Fakeddit* [Kai Nakamura 2019]: *Dataset* composto por 628.501 notícias rotuladas como *Fake* e 527.049 rotuladas como *True*. Estas notícias contém o texto, imagem e os dados dos comentários;

D15) *FakeNewsAMT* [Pérez-Rosas et al. 2018]: As notícias falsas e legítimas são fornecidas em duas pastas separadas. Cada pasta contém 40 notícias de seis domínios diferentes: tecnologia, educação, negócios, esportes, política e entretenimento;

D16) *FakeNewsData1* [Horne and Adali 2017]: São dois *datasets* onde o primeiro contém notícias rotuladas como *Fake e Real* retiradas a partir do *BuzzFeed*. Já o segundo contém notícias políticas rotuladas como *Real, Fake e Sátira* obtidas, randomicamente, durante as eleições americanas de 2016;

D17) *FakeNewsNet1* [Shu et al. 2017] [Shu et al. 2019b] [Sharma et al. 2019] [Gupta et al. 2018] [Shu et al. 2019a]: Esta base de dados, coletada do Twitter, fornece 211 notícias *Fake* e 211 notícias *Real*, rotuladas a partir do *BuzzFeed* e *PolitiFact*. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

¹⁷<https://www.similarweb.com/top-websites/category/News-and-media>

¹⁸<https://en.wikipedia.org/wiki/List-of-fake-News-websites>

D18) *FakeNewsNet2* [Shu et al.][Sharma et al. 2019]: Esta base de dados, coletada do Twitter, fornece 6.480 notícias *Fake* e 17.441 notícias *Real*, rotuladas a partir do *GossipCop*¹⁹ e *PolitiFact*. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

D19) *FakeNewsSet* [da Silva et al. 2020]: Repositório, coletado do Twitter, contendo 300 notícias *Fake* e 300 notícias *Not Fake* em português, rotuladas a partir de agências de checagem de fatos nacionais. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

D20) *Kaggle*²⁰: Este conjunto de dados contém texto e metadados de 244 sites, totalizando 12.999 postagens. Os dados foram extraídos usando a *API webhose.io*. Cada site foi rotulado de acordo com o *BS Detector*, sendo que as fontes de dados sem rótulo foram categorizadas como *Bs*;

D21) *KV* [Dong et al. 2014]: Nesta base as notícias têm sujeito, predicado e objeto. Cada notícia tem um rótulo que indica a probabilidade da mesma ser verdadeira. A ferramenta, por meio de uma fusão de conhecimentos, cria um grafo relacionando o sujeito com o objeto para medir a quantidade de interações e, assim, gerar automaticamente o *dataset*;

D22) *Large-scale Training Dataset e Small-scale Evaluation Dataset* [Helmstetter and Paulheim 2018]: No *Large-scale Training Dataset* cada tweet de uma fonte confiável é rotulado como notícia real e cada tweet de uma fonte não confiável é rotulado como uma *Fake News*. As 46 fontes confiáveis e 65 não confiáveis foram obtidas através de pesquisas em sites e os *tweets* foram coletados a partir dessas fontes. No total, foram coletados 401.414 exemplos, nos quais 110.787 foram rotulados como *Fake News*, enquanto 290.627 foram rotulados como *Real News*. O *Small-scale Evaluation Dataset* contém 116 *tweets* rotulados manualmente a partir do *PolitiFact*;

D23) *LIAR* [Wang 2017]: Esta base de dados é coletada do *PolitiFact*. Ele inclui 12.836 notícias rotuladas manualmente como *Pants-fire*, *False*, *Barely-true*, *Half-true*, *Mostly true* e *True*. Cabe salientar que os dados referentes ao usuário se resumem ao nome do autor da postagem;

D24) *PoliticalNews* [Castelo et al. 2019]: Para criar o dataset foram usados os sites *Politifact*, *BuzzFeed*, *OpenSources.co* e *Alexa's top 500 news*²¹. O resultado foi um *dataset* com 14.240 páginas de notícias sendo 7.136 páginas vindas de 79 sites não confiáveis e 7.104 vindos de 58 sites confiáveis;

D25) *PolitiFact para XFake* [Yang et al. 2019]: Repositório criado a partir do site *PolitiFact* com 5.104 notícias contendo os atributos *Subject*, *Context*, *Speaker*, *Targeting* e *Statement*. As notícias foram rotuladas como *True* e *False*;

D26) *RST-SVM Dataset* [Rubin et al. 2015]: Esta base de dados foi criada a partir de codificadores, usando notícias do *Bluff the Listener*²². Esse repositório consiste de 144

¹⁹<https://www.gossipcop.com/>

²⁰<https://www.kaggle.com/datasets>

²¹<https://www.alexa.com/topsites/category/News>

²²<https://www.npr.org/bluff-the-listener>

notícias selecionadas, aleatoriamente, de 2010 até 2014;

D27) *Signal Media para Evaluating Machine Learning Algorithms for Fake News Detection* [Gilda 2017]: *Dataset* rotulado com *Fake* ou *Not Fake* criado a partir de uma base de notícias da *Signal Media* e uma lista do repositório de confiança de fontes *Open-Sources.co*. Esse *dataset* contém 11.051 artigos, sendo 3.217 rotulados com falsos;

D28) *Soc-LiveJournal* [Srivastava et al. 2018]: Este repositório não rotulado contém uma rede de relacionamentos formada por 4.847.571 nós e 68.475.391 arestas;

D29) *Twitter e Sina Weibo para CSI* [Ruchansky et al. 2017]: *Dataset* criado com 2.811 artigos rotulados como *Fake* e 2.845 como *True*. A citada base de dados foi obtida a partir do repositório, para detecção de rumores, gerado no artigo [Ma et al. 2016];

D30) *Twitter e Sina Weibo para EANN* [Ruchansky et al. 2017]: A base de dados foi criada a partir de dois *datasets* não originários de *Fake News*. O primeiro repositório foi obtido a partir do *Sina Weibo* contendo 4.749 notícias com rótulos adaptados para *Fake* e 4.779 para *Real*, além de 9.528 imagens. O segundo repositório foi obtido a partir do *Twitter* contendo 7.898 notícias com rótulos adaptados para *fake* e 6.026 para real, além de 514 imagens;

D31) *Twitter e Sina Weibo para Early Detection Through Propagation Path* [Liu and BrookWu 2018]: Este repositório foi criado a partir de três *datasets* usados para detecção de rumores. O primeiro, oriundo da rede social *Weibo*, com os rótulos *rumor (fake)* e *otherwise (true)*. Já os outros dois *datasets*, obtidos do *Twitter*, são rotulados como *fake*, *true*, *unverified* e *non-rumor (debunking of fake)*. As características dos usuários foram obtidas por meio de pesquisas realizadas nas respectivas redes sociais.

D32) *Twitter e Sina Weibo para TCNN-URG* [Qian et al. 2018]: Base de dados que utilizou dois *datasets*. O primeiro *dataset* foi obtido, automaticamente, a partir do *Sina Weibo*. Já o segundo *dataset* foi gerado por um processo manual de coleta de dados. Para tal, foram selecionadas notícias em sites avaliados como confiáveis (*The Guardian*²³) e notoriamente falsos. Com as URLs de todas as notícias coletadas, pesquisas foram realizadas no *Twitter* para cada uma das notícias rotuladas como falsas ou reais.

D33) *Twitter para Automatically Identifying Fake News* [Buntain and Golbeck 2017]: Base de dados que utilizou os *datasets* PHEME (rumor no *Twitter*), CredBank (credibilidade no *Twitter*) e *BuzzFeed News Fact-Checking Dataset* (Checagem de fatos no *Facebook*). Os três *datasets* precisaram ser alinhados com as mesmas características e rótulos;

D34) *Twitter para Content Polluters* [Nasim et al. 2018]: Repositório de dados criado para detecção de *bots*. Esse *dataset*, obtido a partir do *Twitter*, foi rotulado manualmente como *Bot* ou *Não Bot*;

D35) *Twitter para Mitigation via Point Process* [Farajtabar et al. 2017]: Este trabalho realizou experimentos com contas reais no *Twitter* e com uma base de dados sintética, onde foi assumido que 20 nós criaram *Fake News* e outros 20 nós divulgaram notícias verdadeiras;

²³<https://www.theguardian.com/>

D36) *Twitter para TraceMiner* [Wu and Liu 2018]: Conjunto de dados gerado pela coleta de informações do *Twitter* com rotulação a partir do site de checagem de fatos *Snopes*²⁴. Nessa base, os rótulos atribuídos são *Real news* ou *Fake news*;

D37) *Twitter Trec* [Srivastava et al. 2018]: Conjunto de dados gerado pela coleta de informações do *Twitter*, sem rotulação, contendo uma rede de relacionamentos formada por 3.919.215 nós e 5.399.949 arestas;

D38) *Twitter Using One-Class* [Faustini and Covões 2019]: Conjunto de dados gerado pela coleta de informações do *Twitter* e de sites de checagem de fatos contendo 4.392 notícias "fake" e 4.589 notícias "real";

D39) *WhatsApp Using One-Class* [Faustini and Covões 2019]: Conjunto de dados gerado pela coleta de informações do aplicativo de troca de mensagens *WhatsApp* e de sites de checagem de fatos contendo 165 notícias "fake" e 12 notícias "real".

Tabela 1.4: Trabalhos x Datasets

Id	DataSet
T1	DataSet Emergent (D7) com Benchmark NLI: Stanford Natural Language Inference (SNLI)
T2	FakeBr (D11)
T3	FakeNewsAMT (D15), Celebrity (D5) e PoliticalNews (D24)
T4	FakeNewsAMT (D15) e Celebrity (D5)
T5	BuzzFeedNews (2016-10-facebookfact-check modificado) (D3)
T6	Twitter para Automatically Identifying Fake News (D33)
T7	Twitter e Sina Weibo para EANN (D30)
T8	FakeNewsNet1 (D17)
T9	FakeNewsNet1 (D17)
T10	DataSet Emergent (D7)
T11	Twitter e Sina Weibo para CSI (D29)
T12	FakeBr (D11) e FakePedia (D12)
T13	DistrustRank Datasets (D8)
T14	Twitter e Sina Weibo para EANN (D30)
T15	Twitter e Sina Weibo para Early Detection Through Propagation Path (D31)
T16	Signal Media para Evaluating Machine Learning Algorithms for Fake News Detection (D27)
T17	Soc-LiveJournal (D28) e Twitter Trec (D37)
T18	FakeBr (D11), FakeNewsNet2 (D18) e FakeNewsSet (D19)
T19	Facebook para Detective (D9)
T20	Twitter Using One-Class (D38) e WhatsApp Using One-Class (D39) e FakeBr (D11)
T21	Twitter para Mitigation via Point Process (D35)
T22	FakeNewsNet1 (D17)
T23	Fakeddit (D14)
T24	LIAR (D23)
T25	Twitter e Sina Weibo para TCNN-URG (D32)
T26	DataSet Emergent (D7)
T27	Twitter para Content Polluters (D34)
T28	PHEME (dataset para Rumor)
T29	FakeNewsSet (D19)
T30	BuzzFace (D2)
T31	FakeNewsData1 (D16)
T32	RST-SVM Dataset (D27)
T33	Twitter para TraceMiner (D36)
T34	FakeNewsNet2 (D18)
T35	Large-scale Training Dataset e Small-scale Evaluation Dataset (D22)
T36	PolitiFact para XFake (D25)

1.4. Problemas em Aberto

O combate automático às *Fake News* nos MDDN é uma nova e emergente área de pesquisa que, mesmo com estudos já realizados, ainda carece de maior aprofundamento científico. Desta forma, os seguintes problemas são descritos como áreas ainda férteis para o desenvolvimento de novos trabalhos:

- Carência de *datasets* que forneçam, de forma suficiente, os diferentes dados necessários para combater as *Fake News* em diferentes MDDN;
- Trabalhos que levem em consideração aspectos temporais do ciclo de vida da *Fake News* e que, conseqüentemente, possam intervir mais rapidamente;
- Estudos que analisem o aspecto intencional, assim não se limitam a verificar a

²⁴<https://www.snopes.com>

Tabela 1.5: Comparação entre Datasets

Id	Dados									URL
	Publicação			Usuário	Propagação			Usuário	Rede	
	Notícia				Contribuição					
	Texto	Áudio	Imagem		Texto	Áudio	Imagem			
D1	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	https://github.com/thiagovas/bs-detector-dataset				
D2	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	https://github.com/gsanita/BuzzFace
D3	No Dataset		Link para imagem	No Dataset	No Dataset	Link para notícia		Link para notícia	Link para notícia	
D4	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	https://github.com/BuzzFeedNews/2016-10-facebook-fact-check				
D5	No Dataset			No Dataset						http://lit.eecs.umich.edu/downloads.html#undefined
D6	No Dataset			No Dataset				No Dataset	No Dataset	http://compsocial.github.io/CREDBANK-data/
D7	No Dataset			No Dataset						https://github.com/FakeNewsChallenge/fnc-1
D8	Link para notícia	Link para notícia	Link para notícia	No Dataset						
D9				No Dataset				No Dataset		
D10	No Dataset									https://github.com/jghm-f/FACTCK_BR
D11	No Dataset									https://github.com/roneysco/Fake-br-Corpus
D12	No Dataset									http://www.fakepedia.org/
D13	No Dataset	Link para notícia	Link para notícia	No Dataset						https://github.com/jgolbeck/fakenews
D14	No Dataset		No Dataset	No Dataset	No Dataset		No Dataset	No Dataset		https://github.com/entitize/Fakeddit
D15	No Dataset			No Dataset						http://lit.eecs.umich.edu/downloads.html#undefined
D16	No Dataset									https://github.com/BenjaminDHorne/fakenewsdata/blob/master/Horne2017_FakeNewsData.zip
D17	No Dataset		Link para imagem	No Dataset				No Dataset	No Dataset	https://github.com/KaiDMML/FakeNewsNet
D18	No Dataset		Link para notícia	No Dataset				No Dataset	No Dataset	https://github.com/KaiDMML/FakeNewsNet
D19	No Dataset		Link para notícia	No Dataset	No Dataset		No Dataset	No Dataset	No Dataset	https://github.com/kamplus/FakeNewsSetGen
D20	No Dataset		Link para imagem	No Dataset						https://www.kaggle.com/mrissdal/fake-news/data
D21	No Dataset			No Dataset						
D22	No Dataset			No Dataset						http://dws.informatik.uni-mannheim.de/en/research/twitter-fake-news-detection
D23	No Dataset			No Dataset						https://github.com/nishitpate101/Fake_News_Detection/tree/master/liar_dataset ou https://www.cs.ucsb.edu/~william/software.html
D24	No Dataset			No Dataset						https://osf.io/e25q4/
D25	No Dataset			No Dataset						
D26	No Dataset			No Dataset						
D27	No Dataset			No Dataset						
D28								No Dataset		https://snap.stanford.edu/data/soc-LiveJournal1.html
D29	No Dataset			No Dataset	No Dataset			No Dataset	No Dataset	https://github.com/majingCUHK/Rumor_RvNN ou http://alt.qcri.org/~wgao/data/rumdelect.zip
D30	No Dataset		No Dataset	No Dataset						
D31				No Dataset				No Dataset	No Dataset	Twitter 15 e 16 (https://www.dropbox.com/s/7ewzdrbelpmxxu/rumdelect2017.zip?dl=0) e Weibo(http://alt.qcri.org/~wgao/data/rumdelect.zip)
D32	No Dataset			No Dataset	No Dataset					False (https://drive.google.com/open?id=1WRoRVV9j4CSIMFKDwP7DVGaHFJZ445a) e True(https://drive.google.com/open?id=1Jg6W4suN2yWHx65P4QU8HkrrB30MHsu0)
D33	No Dataset			No Dataset				No Dataset	No Dataset	
D34				No Dataset				No Dataset	No Dataset	
D35								No Dataset	No Dataset	
D36				No Dataset				No Dataset	No Dataset	
D37	No Dataset			No Dataset				No Dataset	No Dataset	https://trec.nist.gov/data/tweets/
D38	Link para notícia									https://github.com/phfaustini/BRACIS2019_FAKENEWS
D39	No Dataset									https://github.com/phfaustini/BRACIS2019_FAKENEWS

autenticidade (veracidade) das notícias;

- Extração de características a partir de imagem e/ou áudio, portanto não se limitando as análises de texto (multimodal);
- Métodos que abordem características baseadas na rede que representa a propagação da notícia. Neste caso, inclusive, podem ser aplicadas técnicas baseadas em grafos;
- Pesquisas que, ao invés de realizarem uma classificação binária, utilizem probabilidades e/ou pertinências na detecção. Esta linha de trabalho se baseia no fato de que, normalmente, a *Fake News* é uma mistura de afirmações falsas e verdadeiras;
- Utilização de um comitê de classificadores para determinar se uma notícia é *fake*. Desta forma, pode-se agregar diferentes técnicas de classificação na detecção;
- Utilização de modelos não supervisionados ou semi-supervisionados devido à carência de *datasets* rotulados que possuam variedade de dados;
- Estudo sobre o comportamento distinto da *Fake News* em diferentes comunita-

des (escolar, trabalho e etc) e/ou redes sociais (*Weibo, WhatsApp e etc*). Isto se deve pela possível mudança de comportamento das notícias de acordo com o meio;

- Classificar os usuários de *Fake News* com o objetivo de identificar o seu tipo (*humanos, bots e cyborgs*). Isto se deve pela possível alteração de comportamento das notícias propositalmente falsas de acordo com o seu tipo de usuário divulgador;

- Trabalhos relacionados à intervenção de *Fake News*, tanto para bloqueio quanto para mitigação. Haja vista que o combate às *Fake News* não se limita à detecção, sendo necessária, também, a intervenção sobre a mesma;

- Abordagens que atuem descentralizadas na rede. Esta atuação se destaca, pois quanto mais rápido e extensivo for o combate, menor serão os efeitos nocivos da notícia;

- Abordagens que utilizem o assunto para a análise da notícia, pois assuntos relevantes, normalmente, motivam a criação de notícias intencionalmente falsas;

- Pesquisas que levem em consideração a reputação dos usuários, pois usuários com baixa reputação tendem a ser potenciais divulgadores de *Fake News*;

- Detecção de *Fake News* por meio de recursos da Web Semântica;

- Explicabilidade na Detecção de *Fake News*;

- Aplicação de TransferLearning na Detecção de *Fake News*.

1.5. Considerações Finais

Com a crescente popularidade dos meios digitais de divulgação de notícias (MDDN), cada vez mais pessoas consomem notícias on-line, em vez dos tradicionais meios de comunicação. No entanto, os MDDN também são usados para divulgar notícias intencionalmente falsas, as chamadas *Fake News*, que podem causar fortes impactos negativos. Nesse capítulo foi explorado o combate automático às *Fake News* em MDDN. Para tal, a literatura existente foi revisada objetivando, por meio de um levantamento do estado da arte, fornecer subsídios para pesquisas que busquem desenvolver métodos para o combate automático às *Fake News* em MDDN. Tendo como base essa revisão da literatura, dois aspectos significativos podem ser destacados. O primeiro é a carência de datasets, rotulados com fake e não fake, que disponibilizem não somente os dados da publicação, mas, também, as informações relacionadas à propagação das notícias. O segundo aspecto é que os métodos computacionais, voltadas para a detecção, vêm se adaptando de acordo com as mudanças nas características das *Fake News*. Uma dessas mudanças é a maior similaridade nas características de escrita, presentes no texto, entre as notícias *fake* e não *fake*. Portanto, os métodos que não utilizam somente o conteúdo da notícia na tarefa de detecção de *Fake News* têm apresentado resultados promissores. Nesse grupo particular de métodos, aqueles baseados na reputação dos usuários dos MDDN se apresentam como uma alternativa para a detecção de notícias intencionalmente falsas.

Referências

[Ajao et al. 2019] Ajao, O., Bhowmik, D., and Zargari, S. (2019). Sentiment aware fake news detection on online social networks. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2507–2511.

- [Bhatt et al. 2018] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., and Mittal, A. (2018). Combining neural, statistical and external features for fake news stance identification. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1353–1357, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.
- [Braz and Goldschmidt 2017] Braz, P. and Goldschmidt, R. (2017). Um método para detecção de bots sociais baseado em redes neurais convolucionais aplicadas em mensagens textuais. In SBSeg 2017, pages 501–508. 10/11/2017.
- [Buntain and Golbeck 2017] Buntain, C. and Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In 2017 IEEE International Con on Smart Cloud (SmartCloud), pages 208–215.
- [Campan et al. 2017] Campan, A., Cuzzocrea, A., and Truta, T. M. (2017). Fighting fake news spread in online social networks: Actual trends and future research directions. In 2017 IEEE International Con on Big Data (Big Data), pages 4453–4457.
- [Castelo et al. 2019] Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., and Freire, J. (2019). A topic-agnostic approach for identifying fake news pages. In Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, pages 975–980, New York, NY, USA. ACM.
- [Cazalens et al. 2018] Cazalens, S., Lamarre, P., Leblay, J., Manolescu, I., and Tannier, X. (2018). A content management perspective on fact-checking. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 565–574, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.
- [Charles and Oliveira 2018] Charles, A. and Oliveira, J. (2018). Checking fake news on web browsers: an approach using collaborative datasets.
- [Ciampaglia et al. 2015] Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. PLOS ONE, 1:1–13.
- [Conroy et al. 2015] Conroy, N., Rubin, V., and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Association for Information Science and Technology, 52:1–4.
- [da Silva et al. 2020] da Silva, F. R. M., Freire, P. M. S., de Souza, M. P., de A. B. Ple-namente, G., and Goldschmidt, R. R. (2020). Fakenewssetgen: A process to build datasets that support comparison among fake news detection methods. WebMedia '20, page 241–248, New York, NY, USA. Association for Computing Machinery.
- [de Souza et al. 2020] de Souza, M. P., da Silva, F. R. M., Freire, P. M. S., and Goldschmidt, R. R. (2020). A linguistic-based method that combines polarity, emotion and grammatical characteristics to detect fake news in portuguese. In Proceedings of the Brazilian Symposium on Multimedia and the Web, WebMedia '20, page 217–224, New York, NY, USA. Association for Computing Machinery.

- [Deng et al. 2014] Deng, S., Huang, L., and Xu, G. (2014). Social network-based service recommendation with trust enhancement. Expert Systems with Applications, 41(18):8075 – 8084.
- [Dong et al. 2014] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In ACM SIGKDD international Con on Knowledge discovery and data mining, pages 601–610.
- [Empirico 1997] Empirico, S. (1997). Hipotiposes pirrônicas livro i (tradução de danilo marcondes). O que nos faz pensar, 9(12):115–122.
- [Farajtabar et al. 2017] Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., and Zha, H. (2017). Fake news mitigation via point process based intervention. In Proceedings of the 34th International Con on Machine Learning - Volume 70, ICML'17, pages 1097–1106. JMLR.org.
- [Faustini and Covões 2019] Faustini, P. and Covões, T. (2019). Fake news detection using one-class classification. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), pages 592–597.
- [Ferrara et al. 2016] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. Commun. ACM, 59(7):96–104.
- [Flintham et al. 2018] Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., and Moran, S. (2018). Falling for fake news: Investigating the consumption of news via social media. In Proceedings of the 2018 CHI Con on Human Factors in Computing Systems, CHI '18, pages 376:1–376:10, New York, NY, USA. ACM.
- [Freire and Goldschmidt 2020] Freire, P. and Goldschmidt, R. (2020). Combatendo fake news nas redes sociais via crowd signals implícitos. In Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, pages 424–435, Porto Alegre, RS, Brasil. SBC.
- [Freire and Goldschmidt 2019] Freire, P. M. S. and Goldschmidt, R. R. (2019). Fake news detection on social media via implicit crowd signals. In Proceedings of the 25th Brazilian Symposium on Multimedia and the Web, WebMedia '19, pages 521–524, New York, NY, USA. ACM.
- [Gilda 2017] Gilda, S. (2017). Evaluating machine learning algorithms for fake news detection. In 2017 IEEE 15th Student Con on Research and Development (SCORED), pages 110–115.
- [Golbeck et al. 2018] Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J. B., Falak, W., Gieringer, C., Graney, J., Hoffman, K. M., Huth, L., Ma, Z., Jha, M., Khan, M., Kori, V., Lewis, E., Mirano, G., Mohn IV, W. T., Mussenden, S., Nelson, T. M., Mcwillie, S., Pant, A., Shetye, P., Shrestha, R., Steinheimer, A., Subramanian, A., and Visnansky, G. (2018). Fake news vs satire: A dataset and analysis. In Proceedings of

the 10th ACM Con on Web Science, WebSci '18, pages 17–21, New York, NY, USA. ACM.

- [Guo et al. 2020] Guo, B., Ding, Y., Yao, L., Liang, Y., and Yu, Z. (2020). The future of false information detection on social media: New perspectives and trends. ACM Comput. Surv., 53(4).
- [Gupta et al. 2018] Gupta, S., Thirukovalluru, R., Sinha, M., and Mannarswamy, S. (2018). Cimtdetect: A community infused matrix-tensor coupled factorization based method for fake news detection. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 278–281.
- [Helmstetter and Paulheim 2018] Helmstetter, S. and Paulheim, H. (2018). Weakly supervised learning for fake news detection on twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 274–277.
- [Hendriks et al. 2015] Hendriks, F., Bubendorfer, K., and Chard, R. (2015). Reputation systems: A survey and taxonomy. Journal of Parallel and Distributed Computing, pages 184–197.
- [Horne and Adali 2017] Horne, B. D. and Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Association for the Advancement of Artificial Intelligence.
- [Janze and Risius 2017] Janze, C. and Risius, M. (2017). Automatic detection of fake news on social media platforms. In PACIS 2017.
- [Kai Nakamura 2019] Kai Nakamura, Sharon Levy, W. Y. W. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. ArXiv.
- [Kim et al. 2018] Kim, J., Tabibian, B., Oh, A., Schölkopf, B., and Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In Proceedings of the Eleventh ACM International Con on Web Search and Data Mining, WSDM '18, pages 324–332, New York, NY, USA. ACM.
- [Kshetri and Voas 2017] Kshetri, N. and Voas, J. (2017). The economics of fake news. IT Professional, 19(06):8–12.
- [Li et al. 2015] Li, Y., Gao, J., Meng, C., Li, Q., Su, L., Zhao, B., Fan, W., and Han, J. (2015). A survey on truth discovery. ACM SIGKDD Explorations Newsletter, 17:1–16.
- [Liu and BrookWu 2018] Liu, Y. and BrookWu, Y. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In AAAI Con on Artificial Intelligence, pages 354–361.
- [Liu and Xu 2016] Liu, Y. and Xu, S. (2016). Detecting rumors through modeling information propagation networks in a social media environment. IEEE Transactions on Computational Social Systems, 3(2):46–62.

- [Ma et al. 2016] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K., and Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In International Joint Con on Artificial Intelligence.
- [Ma et al. 2015] Ma, J., Gao, W., Wei, Z., Lu, Y., and Wong, K.-F. (2015). Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Con on Information and Knowledge Management, CIKM '15, pages 1751–1754, New York, NY, USA. ACM.
- [Maia et al. 2021] Maia, I. M. L., de Souza, M. P., da Silva, F. R. M., Freire, P. M. S., and Goldschmidt, R. R. (2021). A sentiment-based multimodal method to detect fake news. In Proceedings of the Brazilian Symposium on Multimedia and the Web, WebMedia '21, New York, NY, USA. Association for Computing Machinery.
- [Mejova and Kalimeri 2020] Mejova, Y. and Kalimeri, K. (2020). Advertisers jump on coronavirus bandwagon: Politics, news, and business. ArXiv, abs/2003.00923.
- [Monteiro et al. 2018] Monteiro, R. A. et al. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In Computational Processing of the Portuguese Language, pages 324–334, Cham. Springer International.
- [Moraes et al. 2019] Moraes, M. P., de Oliveira Sampaio, J., and Charles, A. C. (2019). Data mining applied in fake news classification through textual patterns. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia '19, page 321–324, New York, NY, USA. Association for Computing Machinery.
- [Moreno and Bressan 2019] Moreno, J. a. and Bressan, G. (2019). Factck.br: A new dataset to study fake news. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia '19, page 525–527, New York, NY, USA. ACM.
- [Mustafaraj and Metaxas 2017] Mustafaraj, E. and Metaxas, P. T. (2017). The fake news spreading plague: was it preventable? In Web Science Con, pages 236–239.
- [Nasim et al. 2018] Nasim, M., Nguyen, A., Lothian, N., Cope, R., and Mitchell, L. (2018). Real-time detection of content polluters in partially observable twitter networks. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1331–1339, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.
- [Passos et al. 2020] Passos, C., Silva, F., Souza, I., Freire, P., and Goldschmidt, R. (2020). Jogos educacionais digitais como ferramentas de apoio à capacitação discente na identificação de fake news escritas em língua portuguesa: Um estudo de caso. In Anais do XXXI Simpósio Brasileiro de Informática na Educação, pages 401–410, Porto Alegre, RS, Brasil. SBC.
- [Passos et al. 2021] Passos, C., Silva, F., Souza, I., Freire, P., and Goldschmidt, R. (2021). Jedi – um jogo educacional digital para apoiar a capacitação discente na identificação de fake news escritas em língua portuguesa: Estudos de caso nos ensinos médio e superior. Revista Brasileira de Informática na Educação, 29(0):634–661.

- [Pérez-Rosas et al. 2018] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2018). Automatic detection of fake news. In International Conference on Computational Linguistics, pages 3391–3401, Santa Fe, New Mexico, USA.
- [Qian et al. 2018] Qian, F., Gong, C., Sharma, K., and Liu, Y. (2018). Neural user response generator: Fake news detection with collective user intelligence. In International Joint Con on Artificial Intelligence, pages 3834–3840.
- [Reis et al. 2019] Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Explainable machine learning for fake news detection. In Proceedings of the 10th ACM Conference on Web Science, WebSci '19, pages 17–26, New York, NY, USA. ACM.
- [Reis et al. 2019] Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Supervised learning for fake news detection. IEEE Intelligent Systems, 34(2):76–81.
- [Rodrigues 2013] Rodrigues, J. G. (2013). O relativismo acerca da verdade refuta-se a si mesmo? Revista Portuguesa de Filosofia, 69(3/4):777–806.
- [Rubin et al. 2015] Rubin, V. L., Conroy, N. J., and Chen, Y. (2015). Towards news verification: Deception detection methods for news discourse.
- [Ruchansky et al. 2017] Ruchansky, N., Seo, S., and Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Con on Information and Knowledge Management, CIKM '17, pages 797–806, New York, NY, USA. ACM.
- [Saikh et al. 2020] Saikh, T., HariPriya, B., Ekbal, A., and Bhattacharyya, P. (2020). A deep transfer learning approach for fake news detection. In International Joint Conference on Neural Networks (IJCNN), pages 1–8.
- [Santia and Williams 2018] Santia, G. C. and Williams, J. R. (2018). Buzzface: A news veracity dataset with facebook user commentary and egos. In AAAI Con on Web and Social Media, pages 531–540.
- [Seo J. 2013] Seo J., Choi S., H. S. (2013). The method of trust and reputation systems based on link prediction and clustering. In IFIP International Con on Trust Management, pages 223–230.
- [Sethi 2017] Sethi, R. J. (2017). Crowdsourcing the verification of fake news and alternative facts. In Proceedings of the 28th ACM Con on Hypertext and Social Media, HT '17, pages 315–316, New York, NY, USA. ACM.
- [Sharma et al. 2019] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Trans. Intell. Syst. Technol., 10(3):21:1–21:42.
- [Sherchan et al. 2013] Sherchan, W., Nepal, S., and Paris, C. (2013). A survey of trust in social networks. ACM Comput. Surv., 45(4):47:1–47:33.

- [Shu et al. 2019a] Shu, K., Mahudeswaran, D., and Liu, H. (2019a). Fakenewstracker: A tool for fake news collection, detection, and visualization. Comput. Math. Organ. Theory, 25(1):60–71.
- [Shu et al.] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 8(3):171–188.
- [Shu et al. 2017] Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. SIGKDD Explor. Newsl., 19(1):22–36.
- [Shu et al. 2019b] Shu, K., Wang, S., and Liu, H. (2019b). Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM '19, pages 312–320, New York, NY, USA. ACM.
- [Souza Freire et al. 2021] Souza Freire, P. M., Matias da Silva, F. R., and Goldschmidt, R. R. (2021). Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. Expert Systems with Applications, 183:115414.
- [Sponholz 2009] Sponholz, L. (2009). O que é mesmo um fato? conceitos e suas conseqüências para o jornalismo. In Revista Galáxia, pages 56–69. PUC-SP.
- [Srivastava et al. 2018] Srivastava, A., Kannan, R., Chelmiss, C., and Prasanna, V. K. (2018). Factcheck: Keeping activation of fake news at check. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18, pages 2079–2081, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- [Testoni et al. 2021] Testoni, G., Souza, M., Freire, P. M., and Goldschmidt, R. (2021). Um método linguístico que combina polaridade, emoção e aspectos gramaticais para detecção de fake news em inglês. In Anais do X Brazilian Workshop on Social Network Analysis and Mining, pages 151–162, Porto Alegre, RS, Brasil. SBC.
- [Tschatschek et al. 2018] Tschatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., and Krause, A. (2018). Fake news detection in social networks via crowd signals. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 517–524, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.
- [UNESCO 2016] UNESCO (2016). Marco de Avaliação Global da Alfabetização Midiática e Informacional (AMI): disposição e competências do país. Cetic.br.
- [UNESCO 2019] UNESCO (2019). Manual para Educação e Treinamento em Jornalismo. UNESCO.
- [Vavilis et al. 2014] Vavilis, S., PetkoviÄž, M., and Zannone, N. (2014). A reference model for reputation systems. Decision Support Systems, 61:147 – 154.

- [Vo and Lee 2018] Vo, N. and Lee, K. (2018). The rise of guardians: Fact-checking url recommendation to combat fake news. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18, pages 275–284, New York, NY, USA. ACM.
- [Vosoughi et al. 2017] Vosoughi, S., Mohsenvand, M. N., and Roy, D. (2017). Rumor gauge: Predicting the veracity of rumors on twitter. ACM Trans. Knowl. Discov. Data, 11(4):50:1–50:36.
- [Wang et al. 2018a] Wang, P., Angarita, R., and Renna, I. (2018a). Is this the era of misinformation yet: Combining social bots and fake news to deceive the masses. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1557–1561, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.
- [Wang 2017] Wang, W. Y. (2017). “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.
- [Wang et al. 2018b] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., and Gao, J. (2018b). Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '18, pages 849–857, New York, NY, USA. ACM.
- [Woloszyn and Nejd1 2018] Woloszyn, V. and Nejd1, W. (2018). Distrustrank: Spotting false news domains. In Proceedings of the 10th ACM Con on Web Science, WebSci '18, pages 221–228, New York, NY, USA. ACM.
- [Wu and Liu 2018] Wu, L. and Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the Eleventh ACM International Con on Web Search and Data Mining, WSDM '18, pages 637–645, New York, NY, USA. ACM.
- [Yang et al. 2019] Yang, F., Pentyala, S. K., Mohseni, S., Du, M., Yuan, H., Linder, R., Ragan, E. D., Ji, S., and Hu, X. B. (2019). Xfake: Explainable fake news detector with visualizations. In The World Wide Web Conference, WWW '19, pages 3600–3604, New York, NY, USA. ACM.
- [Yu et al. 2020] Yu, J., Huang, Q., Zhou, X., and Sha, Y. (2020). Iarnet: An information aggregating and reasoning network over heterogeneous graph for fake news detection. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–9.
- [Zhang et al. 2018] Zhang, Q., Yilmaz, E., and Liang, S. (2018). Ranking-based method for news stance detection. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 41–42, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

- [Zhang et al. 2020] Zhang, T., Wang, D., Chen, H., Zeng, Z., Guo, W., Miao, C., and Cui, L. (2020). Bdann: Bert-based domain adaptation neural network for multi-modal fake news detection. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–8.
- [Zhou and Zafarani 2018] Zhou, X. and Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315.
- [Zhou and Zafarani 2019] Zhou, X. and Zafarani, R. (2019). Fake news detection: An interdisciplinary research. In Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, pages 1292–1292, New York, NY, USA. ACM.
- [Zhou et al. 2019] Zhou, X., Zafarani, R., Shu, K., and Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. In Proceedings of the Twelfth ACM International Con on Web Search and Data Mining, WSDM '19, pages 836–837, New York, NY, USA. ACM.

Capítulo

2

Modelagem da Estratégia e Alinhamento de Recursos Humanos Baseado em Competências

Henrique Prado de Sá Sousa, Eduardo Kinder Almentero

Abstract

The objective of this short course is to present an organizational modeling language that addresses the topic of competency-based human resources strategy and alignment. This theme brings a large volume of concepts, then the short course is practice-oriented, dividing the presentation of the content into well-defined parts, which can also be used separately. The composition of the developed models will foster strategic analysis based on human resources competencies. The models will be developed from a mini-world designed to facilitate learning. In the end, students will have the experience of modeling processes, organizational objectives, context situations, and human resources skills, in a language that integrates these models, allowing analyzes that help in strategic decisions.

Resumo

O objetivo deste minicurso é apresentar uma linguagem de modelagem organizacional que aborda o tema de estratégia e alinhamento de recursos humanos baseado em competências. Como o tema traz um volume grande de conceitos, o minicurso é orientado à prática, dividindo a apresentação do conteúdo em partes bem definidas, as quais podem também ser utilizadas em separado. A composição dos modelos desenvolvidos irá fomentar a análise estratégica baseada em competências de recursos humanos. Os modelos serão desenvolvidos a partir de um minimundo projetado para facilitar o aprendizado. Ao final, os alunos terão a experiência de modelar processos, objetivos organizacionais, situações de contexto e competências de recursos humanos em uma linguagem que integra esses modelos permitindo análises que auxiliam em decisões estratégicas.

2.1. Introdução

2.1.1. Estratégia Organizacional

As organizações experimentaram recentemente o peso de mudanças drásticas ocasionadas por eventos pouco considerados em sua rotina de gestão. A pandemia da

Covid-19 criou circunstâncias diversificadas, resultando em provações difíceis para inúmeras organizações, mas também criando cenários favoráveis para o crescimento e/ou estabelecimento de determinados tipos de negócios.

Apesar de uma pandemia ser um evento com baixa probabilidade de ocorrer no tempo, não é possível determinar com exatidão as condições do porvir. A experiência enfrentada de forma global demonstra que, na prática, os efeitos advindos das relações sociais no mundo são capazes de interferir sobremaneira em variáveis de interesse de muitas organizações, uma vez que afetam, por exemplo, as economias, os custos, consumo, as cadeias de produção, e até o capital humano, por impactos em sua forma costumeira de vida.

As mudanças contínuas no contexto das organizações são um fato. Elas demandam determinado grau de adaptação para que as organizações permaneçam em condições de cumprir os seus propósitos.

Uma forma de minimizar o impacto de mudanças se baseia no estudo de potenciais cenários futuros, antecipando-se no preparo e planejamento de ações para que, no caso de enfrentamento real destes cenários, a organização esteja preparada para se adequar às novas condições, as quais podem demandar alterações impactantes em tempo diminuto.

As estratégias de uma organização são definidas a partir de seus objetivos, que expressam sobremaneira os estados futuros que a organização almeja. Todas as ações que são projetadas na definição do percurso a ser trilhado visando sair de um estado atual para o estado futuro de desejo são, normalmente, pensadas em pormenores para aumentar as possibilidades de mitigar impactos negativos que porventura surjam na caminhada.

Para apoiar esse processo de construção de estratégias, diversas ferramentas foram propostas, por exemplo *Balanced ScoreCard* (BSC) [Kaplan e Norton, 1992], Análise SWOT [Pickton e Wright, 1998], 5 forças de Porter [Porter, 1980], VRIO *Framework* [Barney, 1992] e *Strategyzer canvas* [Strategyzer, 2015].

Neste minicurso apresentamos uma linguagem de modelagem organizacional que permite registrar informações sobre as variáveis que compõem o contexto organizacional, e vinculá-las às estratégias. As estratégias, por sua vez, podem ser descritas em termos de seus objetivos, processos e tarefas¹, permitindo registrar detalhes sobre a sua implementação.

2.1.2. Alinhamento Organizacional

O trabalho minucioso de projeção de estratégias é apenas o início de um ciclo de gestão na organização. Uma vez definidas as estratégias, inicia-se o grande desafio de implementá-las. Este desafio é caracterizado pela tarefa de coordenar a estrutura organizacional para que atenda da melhor forma possível aos interesses da organização.

Como uma organização é composta por muitas áreas com atuações distintas (apesar de complementares no contexto organizacional), existe um vasto campo literário e científico para cada uma dessas áreas. Isso inclui técnicas e métodos específicos para o alinhamento organizacional.

¹Tarefa também é comumente referenciada como Atividade. Neste trabalho adotamos o termo Tarefa.

Também abordamos o conceito de competências, bastante utilizado na gestão de Recursos Humanos (RH). Através da modelagem de competências são registrados os requisitos humanos para a realização de tarefas na organização. A avaliação do alinhamento de RH consiste na medição das capacidades reais de um indivíduo em corresponder aos requisitos das competências necessárias para lidar com as responsabilidades de determinado papel organizacional.

Utilizando a linguagem GPI-HR [Sousa e Leite, 2017a], é possível modelar e relacionar os requisitos de competências ao modelo que descreve estratégias e, através de relacionamentos, manter o rastro de potenciais impactos que eventuais (des) alinhamentos possam criar. Essencialmente, a linguagem auxilia na análise estratégica organizacional e de RH ao apresentar nos modelos marcadores que simplificam o entendimento do cenário existente, demarcado principalmente por informações sobre o “alinhamento de RH” e “Situação de contexto”.

Adotaremos uma abordagem que empregará exemplos simples para facilitar o entendimento da linguagem GPI-HR. Através de uma abordagem “*botton-up*”, apresentaremos todos os modelos envolvidos. Na demonstração de exemplos, apresentamos “minimundos” seguidos de sua respectiva modelagem.

2.1.3. Organização do Capítulo

O minicurso foi organizado da seguinte forma: primeiramente apresentamos brevemente informações sobre a ferramenta de modelagem GPI-HR. Posteriormente, apresentamos o conceito de competências e seus recursos de modelagem. Em seguida abordamos o tema de modelagem de processos, modelagem de estratégia e situações de contexto. Adicionalmente, foram debatidas as formas de uso e propagação de marcadores que são utilizados para auxiliar a análise do alinhamento. Por fim, o uso dos modelos em análises de alinhamento será exemplificado.

Cada tema apresentado é seguido de um exemplo. As experiências e análises dos autores também são apresentadas.

2.2. A Linguagem de Modelagem GPI-HR

Este minicurso apresenta um conjunto de conceitos envolvidos na modelagem da estratégia, competências e alinhamento de RHs, os quais são aplicados na ferramenta GPI-HR¹.

Essa ferramenta é continuamente atualizada para abordar conceitos envolvidos na modelagem organizacional, mais especificamente, orientada ao alinhamento organizacional. Isso representa obter conhecimento de diversas áreas e incorporar à linguagem de modelagem de forma a representar conceitos importantes e seus impactos através dos modelos.

A linguagem GPI (*Goal, Process, Indicators*) foi projetada para permitir maior grau de rastreabilidade entre elementos da camada de processos e elementos da camada de objetivos [Sousa e Leite, 2014], [Sousa e Leite, 2015], visando ampliar a possibilidade de se analisar o alinhamento entre essas camadas. Posteriormente, a linguagem evoluiu para o GPI-HR (*Goal, Process, Indicators – Human Resources*) para melhor abordar os conceitos de estratégia, competência, e alinhamento de RHs [Sousa e Leite, 2017a], [Sousa e Leite, 2017b].

¹gpi.uniriotec.br

A linguagem GPI-HR utiliza “padrões de modelagem” [Sousa e Leite, 2017b] que definem um formato de modelagem para um conceito bem definido na linguagem. Isso permite oferecer modelos para a modelagem de conceitos específicos, bem como a possibilidade de anexá-los como módulos no modelo principal, agregando informações de alinhamento.

Como a linguagem GPI-HR visa, em especial, auxiliar na análise do alinhamento organizacional e do alinhamento da perspectiva de RH, os marcadores se adaptam a expressar estados qualitativos resultantes, eliminando, à primeira vista, a complexidade dos cálculos de indicadores comuns.

Os marcadores são baseados nas cores vermelho, amarelo e verde, que possuem semânticas universais, inclusive usado em sinais de trânsito em todo o mundo. O vermelho representa o PARE, o amarelo representa ATENÇÃO, e o verde representa SIGA.

Se tratando de marcadores que representam indicadores, a analogia que fazemos é que o vermelho representa nível abaixo do esperado, o amarelo em algum nível que represente risco, próximo ao abaixo do esperado (ou próximo do esperado, conforme se desejar definir com ou uso dos marcadores), o verde representa dentro do nível esperado, e adicionalmente incluímos o azul, representando o que houver acima do esperado. A Figura 2.1 ilustra os marcadores e suas cores.



Figura 2.1. Cores dos marcadores no GPI-HR.

As semânticas das cores são adaptadas para representar de forma adequada o elemento em que a marcação estiver relacionada. Entretanto, espera-se que o entendimento rápido da leitura dos modelos por usuários siga por uma semântica genérica, conforme apresenta a Tabela 2.1:

Tabela 2.1. Semânticas abstratas dos marcadores do GPI-HR.

Marca	Possíveis semânticas
	Acima do esperado, melhor caso, contribuição superior, alinhamento superior.
	Satisfatório, esperado, existente, verdadeiro, contribuição positiva, alinhado.
	Abaixo do esperado, pouco insatisfatório, em atenção, em risco, contribuição negativa, pouco desalinhado.
	Muito insatisfatório, insuficiente, inexistente, falso, sem contribuição, muito desalinhado.

Os marcadores são inseridos no topo esquerdo dos elementos mensurados (Figura 2.2).

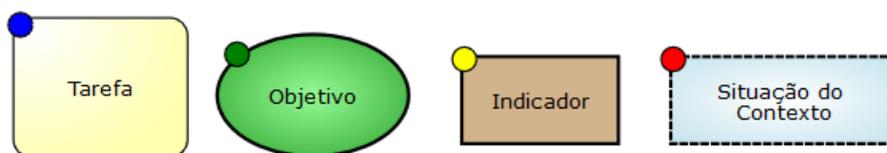


Figura 2.2. Exemplo de marcadores em elementos da linguagem.

No decorrer do minicurso abordaremos os conceitos apresentados são seguidos de exemplo de uso na linguagem GPI-HR. O uso dos marcadores nos diferentes modelos da linguagem é explicado à medida em que esses modelos são apresentados.

2.2.1. Competência

Competência é geralmente¹ definida em termos do nível de “Conhecimento (*Knowledge*), Habilidades (*Skills*) e Aptidões/Talentos (*Abilities*)” que possibilita a uma pessoa o potencial para efetivamente realizar uma tarefa [Curtis *et al.*, 2009], [Stevens, 1994]. A tríade *Knowledge, Skill e Ability* é também conhecida como *KSA Framework* [Tripathi e Agrawal, 2014], em português, referenciado como CHA.

Existem diversas variantes propostas por outros autores [Trichet e Leclère, 2003]. Por exemplo, [Carbone *et al.*, 2005] faz referência a Atitudes (*Attitudes*) no lugar de Aptidões (*Abilities*). Outros autores podem substituir por Comportamento (*behaviour*), Habilidade física (*physical ability*) ou outras variações similares [Klink e Boon, 2002], [Trichet e Leclère, 2003], [Draganidis e Mentzas, 2006]. Independente das variações, trata-se de uma ferramenta bastante difundida na gestão baseada em competências [Draganidis e Mentzas, 2006].

No entanto, essa definição em termos de CHA oferece apenas uma expectativa de bons resultados, uma vez que não é uma regra para o desempenho. O conceito de competência definido por [Fleury e Fleury, 2001], [Dutra, 2007] preenche essa lacuna porque está necessariamente relacionado a um resultado mensurável.

Portanto, espera-se que um indivíduo tenha uma competência determinada somente se tiver os conhecimentos, habilidades e aptidões necessárias e aplicá-las sob determinado desempenho esperado, atendendo as medidas de competências [Fleury e Fleury, 2001], [Carbone *et al.*, 2005], [Dutra, 2007], [Curtis *et al.*, 2009].

Esse conceito se aproxima da definição de [Carbone *et al.*, 2005]: “combinações sinérgicas de conhecimentos, habilidades e atitudes, expressas pelo desempenho profissional dentro de determinado contexto organizacional, que agregam valor a pessoas e organizações”; e da definição do P-CMM, a qual é referenciada como “Competência da força de trabalho” e que “representa uma única integração de conhecimento (*knowledge*), habilidades (*skills*) e aptidões de processos (*process abilities*) adquiridas através da educação especializada e da experiência de trabalho”.

A proposta utilizada neste minicurso considera o conceito de competência em termos de “Conhecimentos, Habilidades e Aptidões”, sendo:

- Conhecimento: é a informação e o entendimento que um indivíduo deve ter para realizar uma tarefa com sucesso.
- Habilidades: são as habilidades técnicas que um indivíduo adquire através do aprendizado, geralmente na aplicação do conhecimento.
- Aptidões: são habilidades/comportamentos individuais, normalmente inatas, porém também podem ser adquiridas.

O conceito de Competência é abordado na literatura de RH considerando diferentes graus de abstração e profundidade (detalhamento). É avaliado a partir de uma perspectiva operacional quando relacionado a pessoas e tarefas de processo [Curtis *et al.*, 2009]; e estratégica, quando vista como um diferencial organizacional [Pralhad e Hamel, 1990]. Porém, também se relaciona com as teorias do aprendizado, tanto no contexto da Psicologia quando faz referência ao ser humano, quanto da Administração, quando faz referência ao conhecimento e aprendizado organizacional.

¹O conceito de competência possui variações em sua definição na literatura motivada por diferentes visões, autores e contexto histórico [Brandão e Guimarães, 2001].

Sob a perspectiva do conhecimento geral, a Competência é comumente relacionada ao conhecimento individual, adquirido através da academia, entretanto, sob a perspectiva estratégica de RH, o conceito de Competência é revisto de forma a atender as demandas das organizações complexas que se encontram em constante evolução e adaptação [Fleury e Fleury, 2011].

Na camada operacional, o conceito de Competência se relaciona aos atores organizacionais em ação; do ponto de vista das tarefas, as Competências são listadas em termos de requisitos para que ela seja realizada de forma correta e satisfatória. Nesse sentido, Competência se encontra relacionada à performance que, por conseguinte, se correlaciona à mensuração. [Dutra, 2007], [Vignotto, 2015], [Fleury e Fleury, 2000] destacam a relação da Competência com o resultado, o que possivelmente justifique, através da avaliação do retorno de uma competência, a ligação ao conceito de estratégia organizacional.

A relação da Competência com o recurso humano e as tarefas as quais são de sua responsabilidade é a visão mais reduzida na observação da interação destes elementos, dentro do contexto organizacional. Novas competências são alcançadas quando as pessoas e suas competências individuais são postas em ação em grupos multidisciplinares. Ou seja, uma equipe de indivíduos é capaz de gerar novas competências através do esforço conjunto de suas competências individuais, principalmente em ambientes multidisciplinares.

Portanto, as competências possuem uma condição especial de se estabelecerem a partir da colaboração de duas ou mais pessoas, entregando suas competências individuais. Por exemplo, na construção de um prédio, vários profissionais atuam juntos, integrando diversas competências que, ao final, alcança a competência “Construção de prédio”. Entretanto, nenhuma das pessoas possui essa competência quando avaliadas em termo de indivíduo. O mesmo efeito se estabelece na definição das “competências organizacionais”.

O termo “entrega de competência” vem da tradução do termo “*competence delivery*”, que é um termo comumente utilizado. Utilizaremos este termo aqui. Esse termo reforça o entendimento da competência como um produto ou serviço bem definido que é entregue a um interessado.

2.2.2. Relações Entre Competências Humanas e as Competências da Organização

A definição das competências organizacionais pode ser direcionada a partir da análise da missão, objetivos estratégicos e valores da organização. As competências individuais e suas configurações em grupos de trabalho podem ser mapeadas a partir das competências organizacionais [Maia, 2011]. No entanto, o inverso também é verdadeiro, quando as competências organizacionais são projetadas a partir das competências humanas existentes.

Pelo seu nível de abstração e complexidade, as competências organizacionais podem expressar os potenciais produtivos da organização, os quais estabelecem suas vantagens competitivas e são de relevância fundamental para alcance dos objetivos estratégicos [Fleury e Fleury, 2001], [Prahalad e Hamel, 1990].

Quando observadas do ponto de vista estratégico, nem todas as competências organizacionais são necessariamente implementadas pela organização. A seleção das competências chave serão balizadas pela estratégia da organização. O conjunto de competências com vínculo estratégico é denominado Competências Essenciais.

Segundo [Prahalad e Hamel, 1990], as competências essenciais (*core competences*) devem ter as seguintes características: oferecer reais benefícios aos consumidores, serem difíceis de imitar e possibilitar acesso a diferentes mercados. São essas competências que diferenciam a organização e que lhe garantem uma vantagem competitiva sustentável perante as demais organizações [Fleury e Fleury, 2004].

As competências essenciais não devem ser confundidas com as competências individuais ou de grupo (ou da força de trabalho, conforme [Curtis *et al.*, 2009] uma vez que se encontram em diferentes níveis de abstração. As competências individuais/grupo estão presentes em um nível inferior (ou mais operacional), relacionadas com o perfil de conhecimento, habilidades e aptidões que são demandadas para se realizar as tarefas organizacionais, funcionando como arcabouço estratégico das competências essenciais da organização [Curtis *et al.*, 2009].

Fleury e Fleury (2004) afirmam que as Competências Organizacionais e Essenciais são como elementos constituídos a partir da combinação de recursos e de múltiplas competências individuais, de tal forma que o resultado total é maior do que a soma das competências individuais.

Na prática organizacional, observamos que a construção das competências organizacionais é decorrente dos processos de colaboração entre competências específicas. Dentro da estrutura organizacional, alguns componentes podem ser provedores de competência. O mais comum são as competências provenientes dos RHs, entretanto os aparatos tecnológicos como softwares e máquinas são introduzidos nas organizações com o objetivo de apoiar na entrega de competências ou ainda, se responsabilizar integralmente por elas.

A tecnologia comumente absorve as competências humanas, mas, não raro, as potencializa de forma substancial, por exemplo, nas grandes máquinas empregadas em processo agrícolas, ou em computadores que analisam grandes volumes de dados.

Independente dos diferentes níveis de qualidade de competência, é através dos processos colaborativos que as competências interagem, gerando competências de valor altamente agregado, não se confundindo com a mera soma de competências (Figura 2.3).

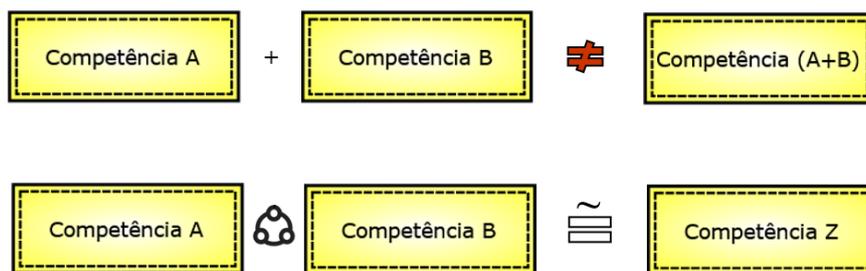


Figura 2.3. Interação colaborativa entre competências.

As competências organizacionais são geralmente apresentadas no portfólio organizacional, caracterizados pelos produtos e serviços que ela oferece aos seus clientes. São as competências de maior valor agregado na organização, e são alcançadas através da contribuição de diversas competências que se agregam através dos processos de produção da organização, os quais demandam de seus responsáveis sua colaboração no processo de construção destas competências finais (Figura 2.4).

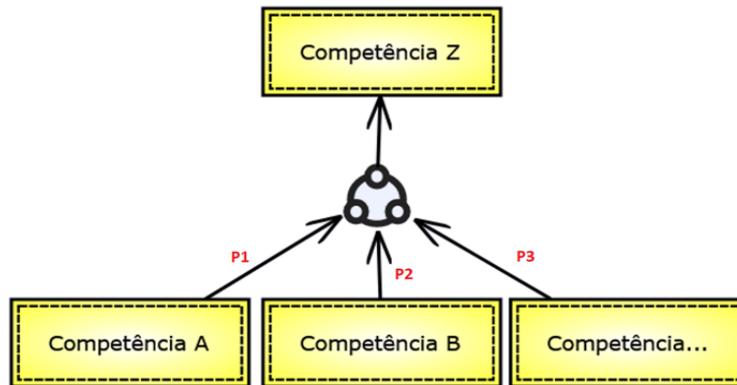


Figura 2.4. Diferentes contribuições de competências no processo iterativo.

Cabe destacar que essas competências entregues pela organização podem servir de insumo para outras organizações agregarem ainda mais valor, gerando entregáveis mais agregados. Um exemplo é a construção de um veículo o qual é composto por partes provenientes de diversas empresas, contendo produtos bem delineados tecnologicamente por organizações especializadas.

Uma estrutura organizacional não reproduz toda a sua capacidade nos produtos e serviços presentes no portfólio, assim como uma pessoa, ocupando determinado papel, não oferece toda a sua capacidade na implementação de suas responsabilidades.

O conhecimento dos potenciais de uma estrutura organizacional é muito importante no contexto estratégico porque apresenta os seus limites, auxiliando nas decisões. Esta é uma das motivações da implementação da gestão por competências. Primeiro porque mapeia os requisitos de competências (mapa de competências) para a implementação de estratégias. Segundo porque mapeia a capacidade de entrega de competências da estrutura organizacional. Sabe-se que somente se gerencia algo que se conhece bem. A gestão de competências torna-se um recurso que apoia a aquisição deste conhecimento, bem como o seu acompanhamento.

Com este conhecimento, decisões estratégicas podem ser mais efetivas, pois é possível pontuar com maior precisão quais caminhos estratégicos a organização é capaz de atender. Ademais, caso se identifique a necessidade de competências que a organização não possui, é possível analisar estratégias para a aquisição de novas competências. Adquirir competências pode, por exemplo, se resumir a uma terceirização, ou ainda, no treinamento de pessoas ou aquisição de insumos tecnológicos.

Para se avaliar as competências de RHs é necessário primeiro mapear as competências requeridas. Essas competências estão relacionadas às responsabilidades de um ator organizacional. Mais detalhadamente, nos processos os quais aquele determinado ator deve implementar.

As competências são requisitos à estrutura da organização que deseja alcançar determinados objetivos. É através do emprego destas competências que todo o processo de realização pode ser implementado dentro dos quesitos de qualidade necessários. Tudo o que se faz demanda alguma competência. É possível observar, sob esta ótica, que desde o nascimento de um indivíduo, competências são desenvolvidas, a começar pelas primordiais que o habilitarão a realizar tarefas mais complexas. Portanto, competências podem ser observadas como requisitos de outras competências de maior valor agregado.

As competências podem ser vinculadas como requisitos a objetivos, processo e tarefas, os quais são elementos fundamentais na modelagem organizacional, e estão relacionados sob uma hierarquia de abstração (Figura 2.5). Um processo é composto por tarefas e é projetado para alcançar determinados objetivos. É possível identificar requisitos de competência para tarefas simples, e de forma agregada, observa-se os requisitos de processos e de seus objetivos.

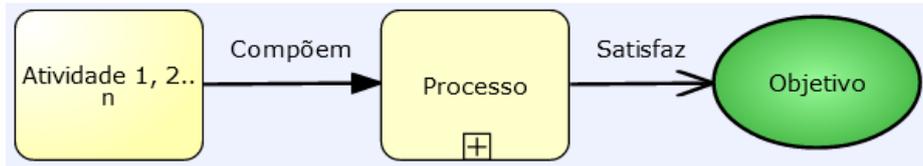


Figura 2.5. Relação entre tarefa x processo x objetivo.

Sob a visão da gestão de competências, essa relação se altera para incorporar o conceito de competências, com sua dinâmica e viés estratégico. A Visão Baseada em Recursos ou *Resource Based View (RBV)* [Penrose, 1959] apresenta uma visão mais maleável e adaptável quando considera que os recursos devem estar aptos a se reorganizarem para oferecer as mesmas competências, ainda que sob o desafio de cenários diversificados. Isso torna a competência um elemento fixo capaz de satisfazer os objetivos para os quais ele foi definido, mas variável nos processos de implementação (Figura 2.6), que se adaptam estrategicamente às mudanças.

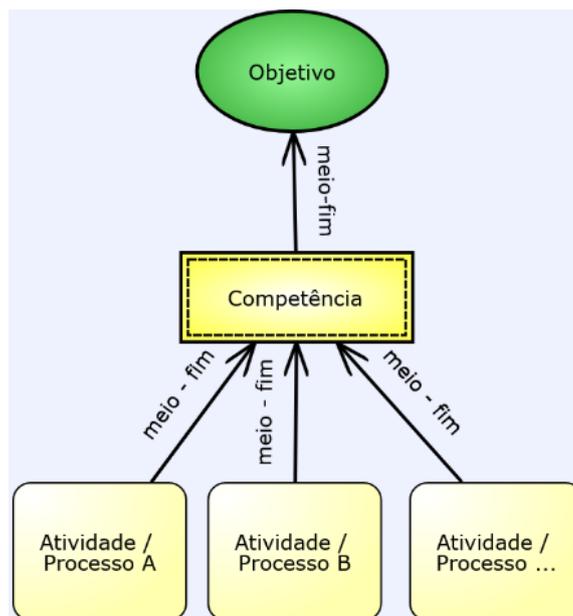


Figura 2.6. Representação da variabilidade de operacionalizações para uma competência através dos relacionamentos de meios-fim.

Essa visão se alinha a necessidade de se planejar ações estratégicas capazes de lidar com as mudanças presentes no mundo, bem como pontua a responsabilidade de se entregar as competências à estrutura da organização. Essa noção é reforçada pela relação de competências vinculadas aos objetivos estratégicos (*core competences*), também definido como o portfólio estratégico de competências da organização [Prahalad e Hamel, 1990].

Também baseado na teoria da RBV [Stirna *et al.*, 2012], [Loucopoulos *et al.*, 2015], o conceito de Capacidade é referenciado por diversas propostas sob as mesmas características identificadas para competência. Inclusive, há autores que adotam o termo “Capacidades principais” (*Core capabilities*) [Azevedo *et al.*, 2015], [Grant, 1996].

2.2.3. Padrão de Perfil de Competências

O perfil de competências segue um padrão de modelagem na linguagem GPI-HR. É a composição do perfil de Conhecimento, Habilidades e Aptidões, conforme apresentado na sessão 2.3.1.

Cada elemento CHA modelado pode se relacionar ao respectivo perfil através de dois possíveis relacionamentos. O primeiro é o relacionamento de “contribuição positiva” (Help), o que significa que aquele elemento ajuda a melhorar o perfil de conhecimento. Entretanto, os elementos que se relacionam através da “contribuição positiva” não são obrigatórios ao perfil. São elementos desejáveis e que, se presentes, fomentam um perfil melhor.

A segunda relação é de “E” (AND). Os elementos relacionados pelo “E” são obrigatórios. Cada elementos CHA também podem ser de dois tipos, mensurado ou não mensurado. Os elementos mensurados estão vinculados a indicadores de desempenho que irão expressar o seu grau através de algum mecanismo de medição. Os elementos que não estão vinculados a indicadores podem ser medidos através de avaliações informais.

A figura 2.7 apresenta o padrão de modelagem do Perfil de Competência.

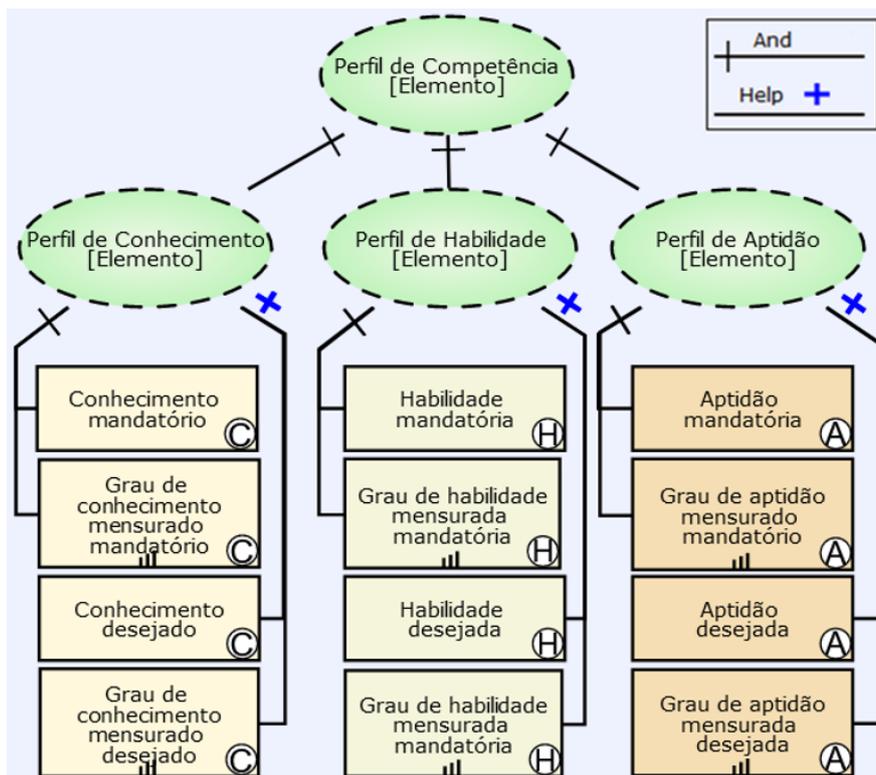


Figura 2.7. Padrão de modelagem para Perfil de Competência.

Este perfil pode ser usado para modelar os requisitos de competência para determinado elemento ou conjunto de elementos, os quais podem ser um papel organizacional, uma competência, um processo, uma tarefa ou um objetivo. Quando o perfil de competência modela requisitos, ele é chamado de Perfil Requerido [Sousa e Leite, 2017a].

O Perfil Requerido pode ser medido, sendo preenchidos por valores nos elementos CHAs, se transformando em um Perfil Real ou Perfil Mensurado. Este perfil pode representar valores de um recurso humano ou grupo de recursos humanos. Entendemos que este perfil também pode ser aplicado à tecnologia, no entanto, não observamos estudos ou relatos deste tipo de aplicação.

A comparação de um Perfil Requerido e um Perfil Real possibilita a avaliação do alinhamento de RH. A partir desta comparação são evidenciados os possíveis desvios de um RH em relação aos quesitos da organização. Portanto, para que seja possível avaliar o alinhamento de RH é necessário primeiramente que a organização entenda bem quais são as suas necessidades nos diversos níveis.

2.2.4. Exemplo - Modelando o Perfil de Competências Requerido

Para exemplificar a modelagem de um perfil de competências, foi utilizado o tema “Fazer bolo” a fim de motivar o pensamento sobre os requisitos de CHA envolvidos neste processo. A Figura 2.8 apresenta o resultado de um perfil de competências requerido elaborado sem formalidades, apenas para compor um exemplo.

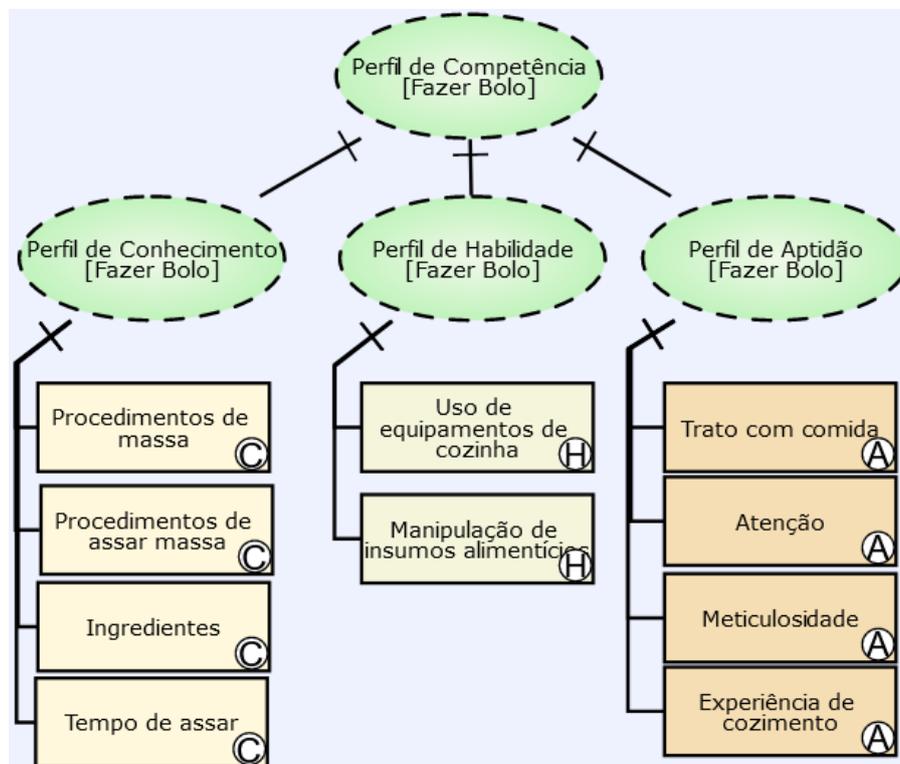


Figura 2.8. Exemplo de perfil de competência requerido para o processo “Fazer bolo”.

Todos os elementos definidos foram relacionados como obrigatórios. Quatro conhecimentos foram mapeados: Procedimentos de massa, representando o conhecimento de como fazer a massa do bolo; Procedimentos de assar, massa representando o conhecimento necessário para colocar a massa para assar corretamente; Ingredientes, representando o conhecimento de todos os elementos que devem compor; e Tempo de assar, representando o conhecimento sobre o tempo adequado de assar determinados tipos de massa de bolo.

No perfil de habilidade foram identificados duas habilidades: uso de equipamentos de cozinha, representando as habilidades necessárias para lidar com a manipulação e operação segura e correta dos equipamentos necessários no processo de

“Fazer bolo”; e Manipulação de insumos alimentícios, representando as habilidades necessárias para operar com alimento, o que permitirá fazer composições adequadas, organização, e limpeza, sem estragar os alimentos ou criar resultados desfavoráveis durante a manipulação dos insumos.

No perfil de aptidões, foram identificados quatro elementos: trato com comida, que representa a aptidão de atuar de forma respeitosa, asseada e responsável com o alimento; Atenção, representando a necessidade de se observar os procedimentos e seus riscos, bem como o tempo de preparo para que não cause acidentes ou estrague o alimento; Meticulosidade, representando a capacidade de observar os pormenores durante o trabalho, e seguir os procedimentos à risca para obter resultados de maior qualidade; Experiência de cozimento, representando a necessidade de já ter experiência nos processos de cozimento, o que permitirá ao indivíduo o melhor trato em circunstâncias inesperadas, contorno de problemas e decisões mais adequadas ao sucesso do trabalho.

Todos esses elementos juntos compõem os requisitos humanos mapeados para que um indivíduo consiga realizar de forma adequada o processo de “Fazer bolo”. Observamos que as aptidões do exemplo são qualidades humanas que auxiliam na qualidade do trabalho, sem as quais, torna-se menor a probabilidade de sucesso. Todos os elementos CHA são obrigatórios, portanto, na ausência de qualquer um deles, ou em alguma avaliação insuficiente, observa-se o risco em algum grau, de não se alcançar a satisfação dos objetivos na implementação do processo, uma vez que se observa a possibilidade de problemas de execução devido a insuficiência de um ou mais CHAs necessários.

2.2.5. Exemplo - Modelando o Perfil de Competências Real

O Perfil Real é o Perfil Requerido quando instanciado para um RH. Nessa instância, todos os elementos CHA são medidos e, ao final, propaga-se os valores para os elementos superiores até exibir o marcador no Perfil de Competências, o qual também representa o marcador de Alinhamento de RH.

Marcadores no Perfil Real / Mensurado

Nos Perfis de CHA são registrados elementos mensuráveis e não mensuráveis.

Os **elementos mensuráveis** recebem a marcação dos seus indicadores de desempenho.

Os **indicadores de desempenho** são usados para medir elementos diversos de forma quantitativa ou qualitativa. Os marcadores expressam o grau de satisfação dos indicadores de desempenho. A classificação semântica dos marcadores dos indicadores é dividida em quatro intervalos, conforme apresenta a Tabela 2.2:

Tabela 2.2. Semântica dos marcadores para Indicadores de desempenho.

Marca	Semântica
	O indicador está acima do intervalo esperado.
	O indicador está dentro do intervalo esperado.
	O indicador está em uma zona próximo do menor valor esperado e/ou pouco abaixo do intervalo esperado, porém ainda aceito. Possibilidade de risco controlado.
	O indicador está abaixo do pior intervalo aceito.

Os *elementos não mensuráveis* representam CHAs de valores binários, ou seja, o RH tem ou não tem o atributo. Para estes casos, a marcação só recebe dois valores, conforme mostra a Tabela 2.3.

Tabela 2.3. Marcas para elementos CHA desejados e mandatórios.

Marca	Semântica
	O recurso humano tem o elemento.
	O recurso humano não tem o elemento.

Marcador de Alinhamento de RH

O marcador do Perfil de Competência representa um marcador de Alinhamento de RH porque consolida os marcadores dos Perfis CHA. Essa consolidação de marcadores é proveniente de uma propagação baseada no pior caso, ou seja, na presença de elementos com diferentes marcadores dentro do perfil, o marcador que o representará será o de menor grau de avaliação. Isso permite a replicação do impacto negativo de possíveis desalinhamentos, o que facilita a sua identificação através dos modelos.

A Tabela 2.4 apresenta as semânticas possíveis para o marcador de Alinhamento de RH.

Tabela 2.4. Marcadores para Alinhamento de RH.

Marca	Semântica
	Competente para executar a tarefa com maestria (tem todas as qualidades necessárias em níveis acima do satisfatório). RH alinhado, desempenho acima do esperado.
	Competente para executar a tarefa (tem todas as qualificações necessárias no nível satisfatório. Algumas qualidades podem estar em níveis acima do satisfatório). RH alinhado, desempenho satisfatório.
	Parcialmente incompetente para executar a tarefa (tem pelo menos uma qualificação necessária em nível de atenção. As outras qualidades estão em níveis satisfatórios ou acima). Desalinhamento de RH, risco assumido.
	Incompetente para executar a tarefa (tem pelo menos uma qualificação necessária em nível abaixo do aceito. As outras qualidades tornam-se indiferentes). Desalinhamento de RH.

A Figura 2.9 apresenta um exemplo de Perfil Real, instanciado do Perfil Requerido presente na Figura 2.8, para a Sra. Mafalda.

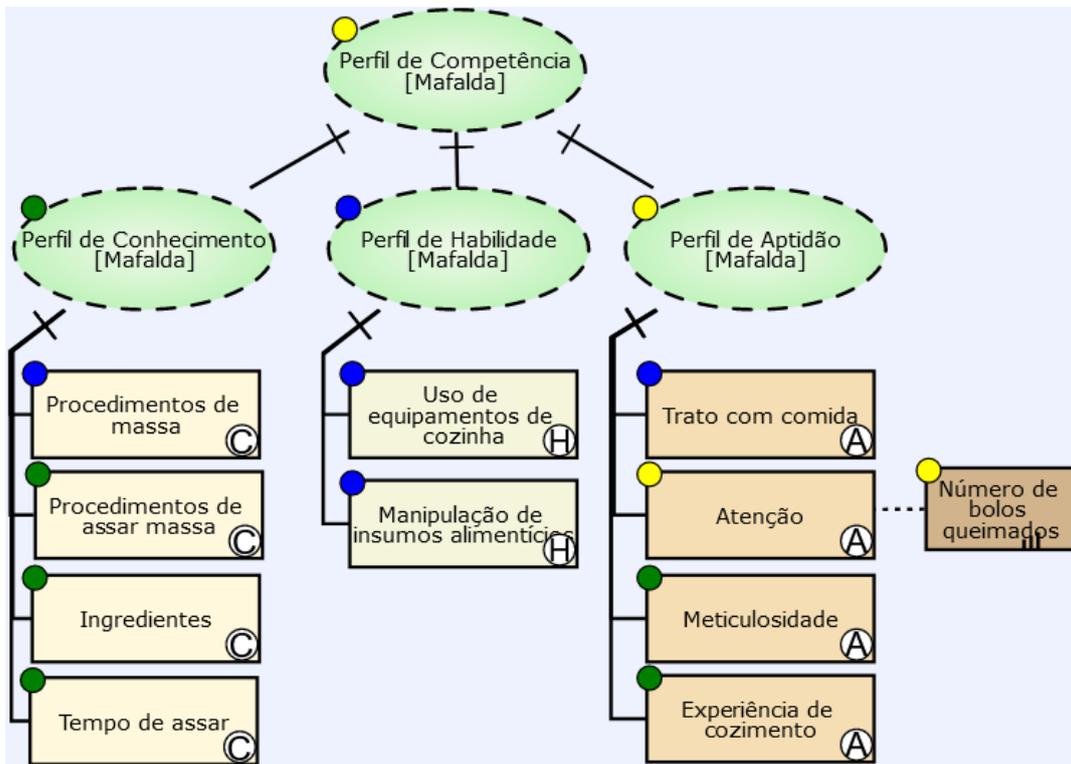


Figura 2.9. Perfil Real da Sra. Mafalda.

O indicador “Número de bolos queimados” demonstra que a Sra. Mafalda não tem apresentado bons resultados, ainda que seja tolerado essa falha. Isso denota (segundo o modelo) falta de atenção. É possível que a Sra. Mafalda esteja desatenta durante o cozimento, o que tem gerado resultados inconvenientes, como a queima do bolo.

Entretanto, nos outros elementos CHA a Sra. Mafalda possui excelente avaliação. Na propagação dos marcadores, o Perfil de conhecimento ficou verde; o Perfil de habilidade, azul; e o Perfil de Aptidão, amarelo. O valor amarelo do Perfil de Aptidão é proveniente da avaliação do indicador, que propagou para a Aptidão Atenção, também propagando para o Perfil de Aptidão, considerando que este é o pior caso neste perfil. Entre os perfis (verde, azul e amarelo), a propagação também foi feita com o marcador amarelo para o Perfil de Competência, sendo esta a avaliação de Alinhamento de RH da Sra. Mafalda.

2.3. Modelagem de Processos

Os processos apresentam uma sequência temporal de tarefas que devem ser realizadas de forma satisfatória para que, ao final da execução do processo, o(s) objetivo(s) o(s) qual(is) ele se propõe(m) alcançar, seja(m) alcançado(s).

A linguagem GPI-HR reutiliza alguns dos elementos de notação da BPMN [OMG, 2011], bem como personaliza outros para representar novos conceitos, por exemplo, a tarefa e processo especializado.

Para exemplificar a modelagem de um processo, utilizaremos uma notação simplificada contendo apenas os conceitos fundamentais, descritos na Tabela 2.5:

Tabela 2.5. Notação simplificada para modelagem de processos.

	<p>Uma tarefa representa uma ação que demanda competências humanas comuns à formação geral relacionada ao papel organizacional.</p>
	<p>Entradas e saídas de tarefas que podem representar elementos físicos, como um produto, ou abstrato, como uma informação.</p>
	<p>Um operador lógico é usado para controlar divergência e convergência nos fluxos sequenciais de um processo.</p>
	<p>Eventos expressam fatos que ocorrem durante a execução de um processo. Pode ser classificado em evento inicial, intermediário e final.</p>

O processo a ser modelado prossegue no tema de “Fazer bolo”, aplicado ao modelo de competências apresentado anteriormente. Utilizamos uma descrição simples de receita de bolo, obtida em uma busca no google¹. Os passos são transcritos a seguir:

1. Bata as claras em neve e reserve.
2. Misture as gemas, a margarina e o açúcar até obter uma massa homogênea.
3. Acrescente o leite e a farinha de trigo aos poucos, sem parar de bater.
4. Por último, adicione as claras em neve e o fermento.
5. Despeje a massa em uma forma grande de furo central untada e enfarinhada.
6. Asse em forno médio 180 °C, preaquecido, por aproximadamente 40 minutos ou ao furar o bolo com um garfo, este saia limpo.

Nosso exercício resultou no modelo apresentado na Figura 2.10.

A modelagem, de uma forma geral, possui desafios intrínsecos, especialmente quanto a completude das informações. Observamos que textualmente o processo de “Fazer **bolo**” é apresentado em 6 linhas. No entanto, ao formalizar o processo através de uma linguagem de modelagem, observa-se que há muitos detalhes implícitos. Este caso apresenta bem a importância do uso do modelo para formalizar os procedimentos e seus pormenores de tal forma que seja obtido o “algoritmo” de realização do processo. Não obstante, a modelagem de processo utiliza elementos como operadores lógicos e eventos para demonstrar as possibilidades e diferentes caminhos que podem ser seguidos dentro de uma instancia de processo.

Por mais simplória que seja a informação escrita no site, uma série de insumos e produtos parciais são administrados durante a execução do processo. A formalização em um modelo auxilia a identificar os detalhes de cada tarefa e, eventualmente, identificar ausência e/ou inconsistências de informações. O entendimento do modelador em relação às informações de origem também pode afetar o modelo final.

¹ Resultado utilizado: <https://www.tudogostoso.com.br/receita/29124-bolo-simples.html>

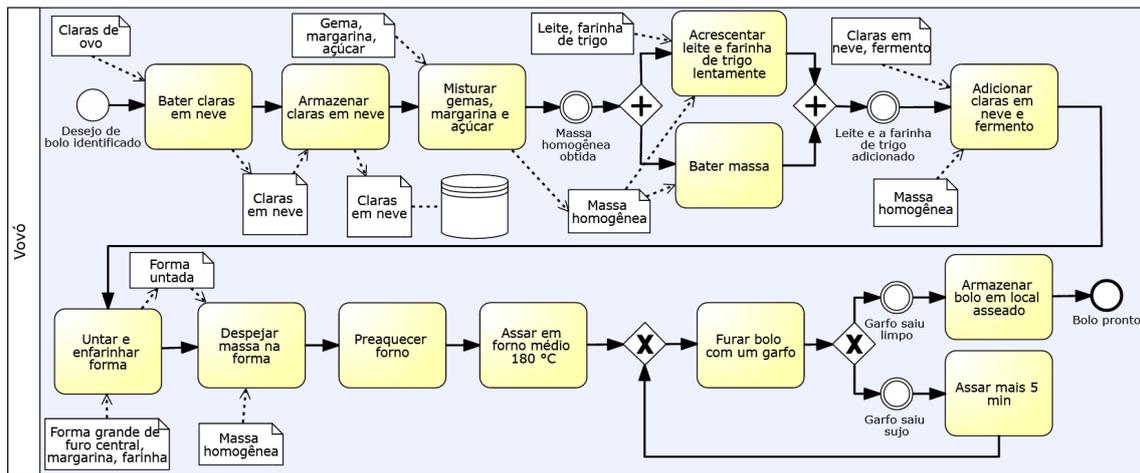


Figura 2.10. Exemplo de processo “Fazer bolo”.

2.3.1. O Perfil Requerido em Diferentes Tarefas/Processos

Um processo é composto por tarefas atômicas que podem realizadas por recursos humanos. Cada tarefa possui o seu grau de complexidade e importância aos objetivos organizacionais, demandando diferentes perfis profissionais para aquele que for realizá-la.

Em geral, papéis organizacionais demandam um perfil profissional específico, o qual se define por diversas competências básicas na sua formação (ex. médico e engenheiro). Entretanto, existem tarefas que podem demandar especializações diversas que se destacam do que é considerado básico. Normalmente são tarefas mais complexas e sensíveis, que oferecem o risco de gerar problemas à execução do processo caso sejam negligenciadas ou realizadas de forma insatisfatória. Os impactos dependem da relação do produto destas tarefas com o resultado do processo, ou ainda, com a importância de seus resultados à indicadores importantes.

Baseado nisso, são usados os conceitos de Tarefas Gerais e Tarefas Especializadas. As Tarefas Gerais demandam perfil de competência similar, intrínseco à todas as tarefas comuns. As Tarefas Especializadas demandam maior refinamento de competências, por exemplo, devido ao seu grau dificuldade, riscos, especialização, importância estratégica. De forma geral, a Tarefa Especializada apresenta algo que justifica a diferenciação na seleção de determinado recurso humano o qual irá se responsabilizar pela tarefa.

A modelagem do Perfil Requerido destes dois tipos de tarefas é necessária para definir os requisitos em que o ator organizacional demandará na expectativa que suas responsabilidades possam ser satisfeitas e, conseqüentemente, seus objetivos serem atingidos.

O Perfil Requerido das Tarefas Gerais será potencialmente suprido pela formação geral demandada pelo ator organizacional. O Perfil Requerido das Tarefas Especializadas é especialmente elaborado para este tipo de tarefa. A expectativa é que cada Tarefa Especializada tenha seu conjunto de requisitos específicos, para que seja avaliada especificadamente na questão do alinhamento de RH.

Por exemplo, para uma Tarefa Especializada de decisão estratégica, o entendimento da estratégia é fundamental, uma vez que uma decisão desfavorável poderia resultar em um desalinhamento de maior risco à organização. Nestes casos, observa-se que o Perfil Requerido da tarefa deve ser modelado registrando como quesito maior nível de indicação do entendimento da estratégia do que no Perfil Requerido das

Tarefas Gerais. Assim, considera-se que o recurso humano que for alocado para a realização de uma Tarefa Especializada deve estar mais qualificado e seja capaz de concretizar a tarefa de forma satisfatória.

A literatura de RH também aborda a definição da complexidade em tarefas e postos de trabalho. Para [Dutra, 2007], a diferenciação de níveis de complexidade permite perceber com maior nitidez o processo de desenvolvimento das pessoas e o levantamento de necessidades de treinamento de maneira mais aproximada da realidade de trabalho e, portanto, de maneira mais precisa. Maia (2011) define o conceito de nível de entrega, que determina a intensidade mínima requerida para um determinado posto de trabalho, o que permite traçar trilhas de aprendizagem de acordo com os seus objetivos de carreira.

Apesar das indicações de alinhamento de RH, nem sempre é possível ter o recurso humano competente disponível. A decisão de alocar um recurso humano com qualificação menor à demandada torna-se um elemento de decisão da organização e um indicativo de risco, ainda que os resultados alcançados sejam satisfatórios.

2.3.2. Exemplo de Tarefa Especializada

Considerando o processo de “Fazer bolo” descrito na Figura 2.10, alteramos, para exemplificar, a tarefa “Assar em forno médio a 180°C” para “Assar massa em forno industrial de lastro” (Figura 2.8).

Ao se inserir esse tipo de tecnologia industrial, obrigatoriamente se faz necessário rever o conjunto de requisitos de competências necessários para implementar este processo já que agora é exigida a capacidade de operar adequadamente um forno industrial do tipo “de lastro”. Portanto, há uma tarefa que demanda um perfil mais específico, com especialidade na operação deste forno. Isso foge ao currículo genérico de pessoas que se propõem a fazer bolo.

Para diferenciar esses processos/tarefas, a notação é alterada (Tabela 2.6 e Figura 2.11) pela cor, e há a expectativa de existir um Perfil Requerido específico para estes processos/tarefas, tanto para identificar os CHAs específicos, mas especialmente para permitir a avaliação de alinhamento de RH em separado dos processos/tarefas gerais.

Tabela 2.6. Notação específica para processos/tarefa complexa/especializada.

	<p>Uma tarefa complexa/especializada representa uma tarefa que demanda competências humanas específicas para que seja realizada de forma satisfatória.</p>
	<p>Um processo complexo/especializado é composto por tarefas que demandam competências humanas específicas para que seja realizada de forma satisfatória.</p>

Esses processos/tarefas especializadas podem representar também partes críticas na organização, as quais devem ser acompanhadas. Por exemplo, tarefas vinculadas a objetivos estratégicos, ou tarefas sensíveis que estão diretamente relacionadas ao desempenho.

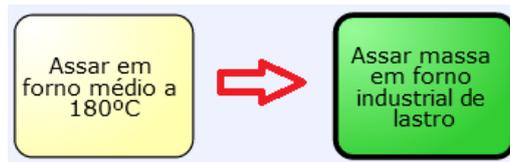


Figura 2.11. Mudança de tipo de tarefa, geral para especializada.

2.3.3. Relacionando Perfis, Tarefas, Processos, Competências e Objetivos

Conforme apresentado na seção 2.2.2, tarefas/processos são um meio de se entregar competências, e as competências são um meio de se satisfazer objetivos (Figura 2.6). Nesta sessão descrevemos como essas relações ocorrerem nos modelos de processos.

Na seção anterior, afirmamos que em um modelo de processo existem transformações nos artefatos que são processados em suas tarefas. Quanto mais complexo um processo, maior a possibilidade que ele tenha produtos parciais. Esses produtos parciais são bem definidos por indivíduos que conhecem o domínio para o qual ele se aplica. Cada trecho de um processo pode representar algo com semântica bem definida, e denotar um ponto em que se pode afirmar que algo foi alcançado. Identificar essas partes é importante para que se entenda melhor o funcionamento do processo, e isso permitirá analisá-lo de forma mais aprofundada.

Todo processo é definido com pelo menos um objetivo. Esse objetivo pode ser decomposto em subobjetivos se for possível também delinear trechos bem definidos de um processo, uma vez que são estes trechos que implementam os subobjetivos. Nos casos em que um único processo é relacionando a dois ou mais objetivos, fica a pergunta: todos os objetivos são alcançados somente ao fim da instância do processo ou eles são alcançados em momentos diferentes durante a execução do processo?

Pontuar cada objetivo e sua relação exata com as tarefas de um processo é um passo importante que irá manter um rastro mais aprofundado, o que consequentemente permitirá identificar as tarefas que estão vinculadas a ele, bem como os atores, artefatos, indicadores e, especialmente na linguagem GPI-HR, a relação com os Perfis de Competência, ou seja, os requisitos de RH.

Para demarcar trechos de um processo, a linguagem GPI-HR utiliza um elemento de agrupamento (*agroupment*) que agrupará os elementos de um processo que são implementados para satisfazer determinado subobjetivo do processo. Seguindo a relação orientada por competências, o trecho irá entregar uma competência de menor valor agregado, porém, bem definida.

A Figura 2.12 apresenta o processo de “Concessão de empréstimo” contendo quatro agrupamentos que são vinculados especificamente as competências que entregam e, por sua vez, as competências se relacionam aos objetivos que satisfazem.

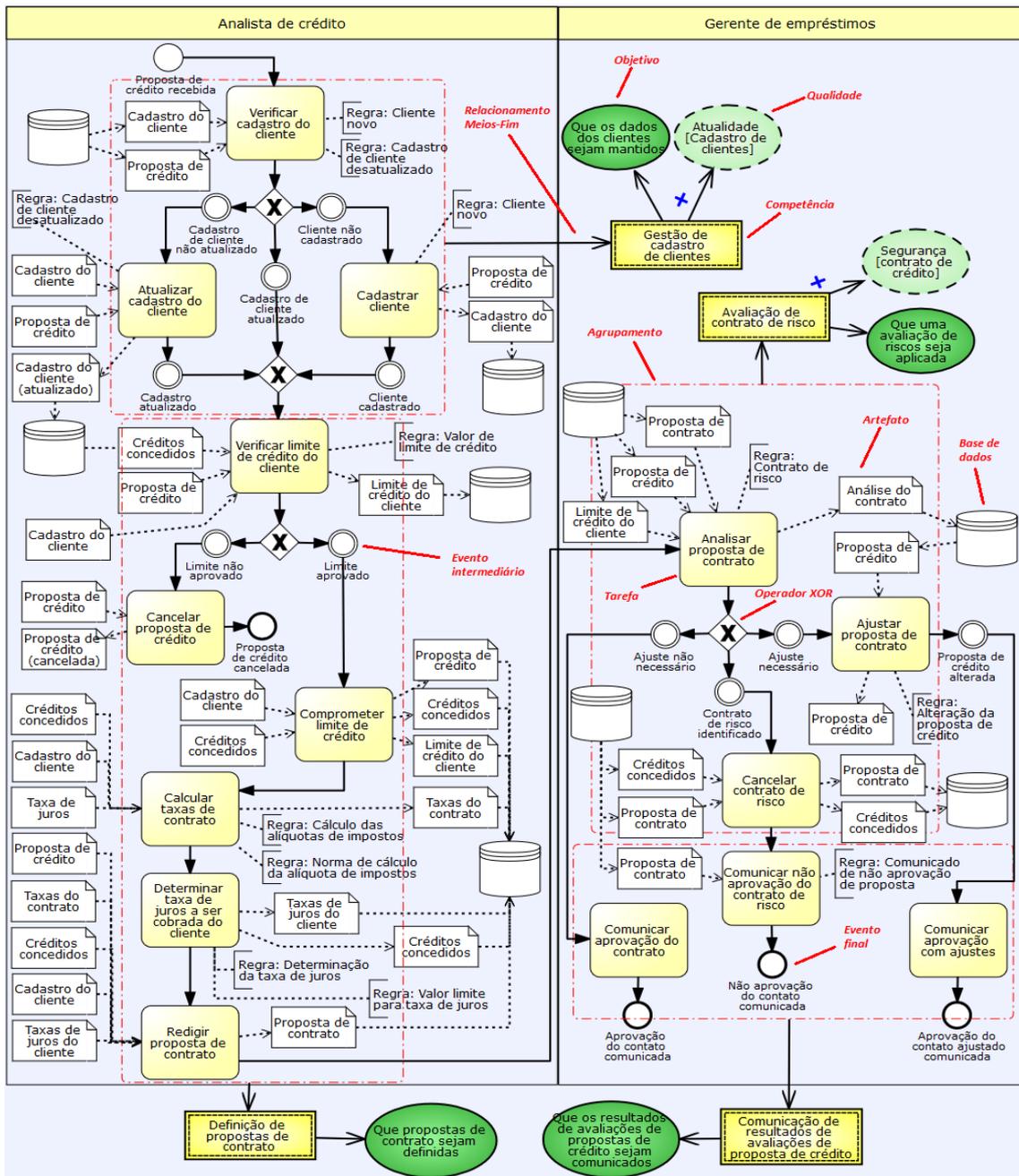


Figura 2.12. Processo “Concessão de empréstimo”.

Este processo mais “robusto” e “detalhado” é apresentado como exemplo mais completo para demonstrar o uso da linguagem GPI-HR no nível de processos.

A “decomposição do processo” permite isolar partes para permitir sua análise. A figura 2.13 apresenta o Perfil de Competência Requerido instanciado apenas para a competência “Definição de propostas de contrato”.

Eventuais problemas na entrega dessa competência podem ser investigados através da sua relação com as tarefas específicas que a implementam, bem como na avaliação do Perfil Real que estará vinculado a esta competência.

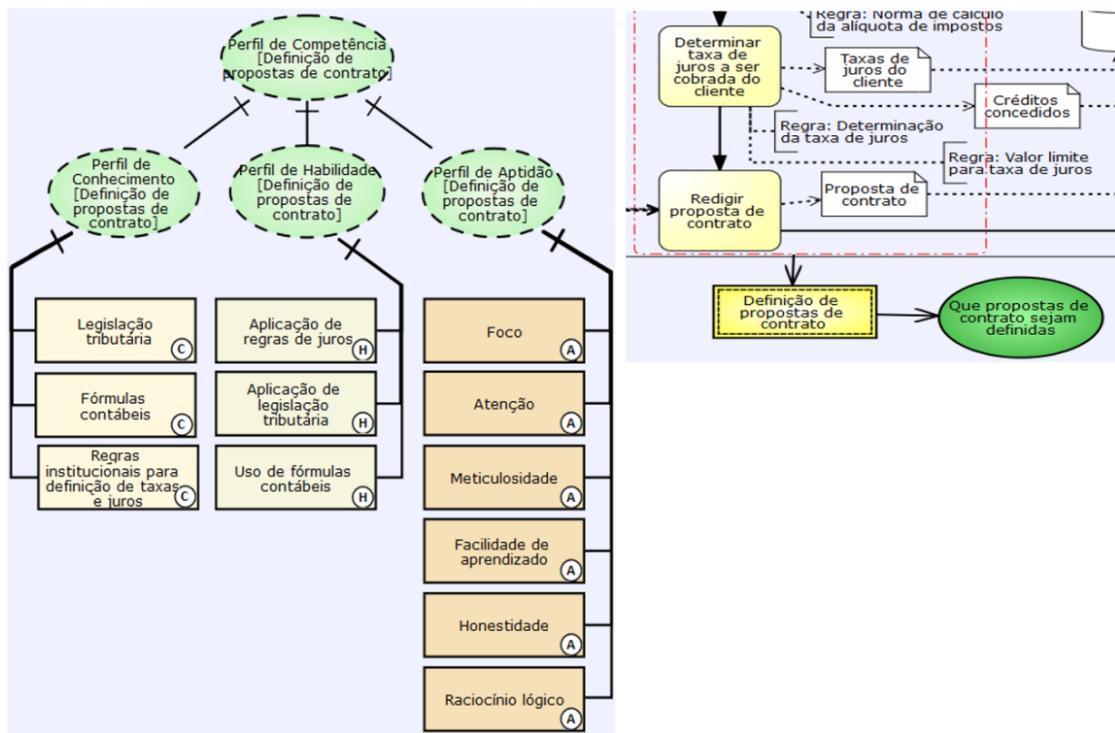


Figura 2.13. Exemplo de processo “Fazer bolo”.

2.4. Estratégia

O conceito de estratégia é discutido na literatura sob diversos aspectos, tornando o conceito amplo e diversificado. Segundo [Mckeown, 2019] “*estratégia é moldar o futuro - é como as pessoas alcançam o fim desejável com os meios disponíveis*”. Entretanto, projetar o futuro significa lidar com incertezas. Projetar cenários hipotéticos está vinculado a planejar ações de forma antecipada, visando lidar da melhor forma possível com as mudanças de contexto.

Segundo a visão apresentada na linguagem i^* por [Yu, 1994], a estratégia é definida pelo raciocínio (*reasoning*) de um ator para se alcançar elementos de uma forma intencional. O raciocínio mencionado por Yu se refere a descrição de um conjunto de elementos intencionais, que delineiam diferentes caminhos, os quais configuram meios de implementação para se alcançar objetivos.

Sob a ótica organizacional, as organizações definem estratégias como uma forma de sobrevivência, uma vez que normalmente se encontram em mercados competitivos. Uma boa estratégia pode permitir que as empresas alcancem posições de vantagem em relação aos seus concorrentes.

2.4.1. A Estratégia e a Camada Estratégica de uma Organização

Organizações são normalmente divididas nas camadas estratégica, tática e operacional. Essa divisão em camadas cria uma linha fictícia que representa uma hierarquia decisória.

A camada estratégica visa delinear os caminhos que a organização deve seguir, o que normalmente são decisões de longo prazo, tomadas por indivíduos do alto escalão de gerência da empresa. A camada tática é intermediária, e normalmente se responsabiliza por decisões de médio prazo, sendo ainda uma camada gerencial que define responsabilidades para a camada operacional. A camada operacional é responsável pela implementação dos processos de produção da organização.

Vulgarmente conhecida como “chão de fábrica”, toma decisões de curto prazo, visando normalmente manter os processos em execução. Pode ter uma parte gerencial, por exemplo, composta de líderes/coordenadores de equipe.

Apesar desta relação hierárquica, em todas as camadas pode (e deve) haver estratégia. Por exemplo, na camada operacional, no setor de vendas, um recurso humano pode implementar as suas estratégias de venda, visando obter bom desempenho. Uma equipe médica pode traçar um plano estratégico de tratamento, visando curar determinada doença.

Portanto, o conceito de estratégia não deve ser confundido com o de camada estratégica. Todos os papéis organizacionais podem implementar estratégias para alcançar seus objetivos.

2.4.2. Modelando a Estratégia

As estratégias são definidas a partir de objetivos centrais. Por exemplo, no nível organizacional, os objetivos estratégicos são definidos a partir dos objetivos primários de qualquer organização chamados de visão e missão, sendo a visão um objetivo de longo prazo, e a missão um objetivo mais imediato. Tudo o que se prossegue destes elementos são decomposições de objetivos em outros objetivos menos abstratos, os quais serão vinculados, em algum momento, aos departamentos como suas responsabilidades. Estes, por sua vez, irão desdobrar suas metas pela estrutura que administram até alcançar objetivos mais concretos, ligados aos elementos que os implementam (por exemplo, tarefas, processos, competências – comumente chamados de *operacionalização*).

Na literatura de modelagem organizacional, a possibilidade de se escolher entre diferentes formas de operacionalização é chamada de “variabilidade” [Liaskos *et al.*, 2007]. A variabilidade é considerada um aspecto positivo porque oferece diferentes possibilidades que possuem características próprias e são melhor adequadas a diferentes circunstâncias.

Neste minicurso, exemplificaremos casos mais simples para facilitar o entendimento. Portanto, iniciaremos a modelagem da estratégia a partir de um objetivo.

Ao se definir um objetivo surgem questões que devem ser respondidas para se traçar possíveis planos capazes de satisfazê-lo. Por exemplo, “Como alcançar esse objetivo?”, “Existem outras formas de satisfazer este objetivo?”, “O que é necessário para se alcançar este objetivo?”, “Os recursos estarão disponíveis?”, “Existem pontos fracos ou ameaças que podem atrapalhar a satisfação deste objetivo?”, “Existem pontos fortes ou oportunidades que poderão apoiar a satisfação deste objetivo”. Estas e inúmeras outras perguntas podem ser feitas na projeção de operacionalizações para o objetivo.

Eventualmente, o conhecimento elencado irá permitir a definição operacionalizações de qualidade, que possuem características distintas que as caracterizarão como opções mais ou menos adequadas a diferentes circunstâncias.

2.4.3. Exemplo – Modelando Possíveis Operacionalizações e Situações de Contexto

Na linguagem GPI-HR existe um relacionamento específico, proveniente do reuso da linguagem *i** que se chama meios-fim (*means-end*) [Yu, 1994].

O relacionamento de meios-fim liga um ou mais elementos primários que representam caminhos possíveis (meio) para se alcançar um elemento secundário, que representa o alvo (fim) desejado. A decisão pela realização do “meio” caracteriza a

intenção de se alcançar o “fim”. Na abordagem orientada a competências, as competências são os meios definidos para se alcançar os objetivos da organização, definidos como fim.

O relacionamento de meios-fim é representado por uma seta simples, com ponta aberta e alongada (Figura 2.14).

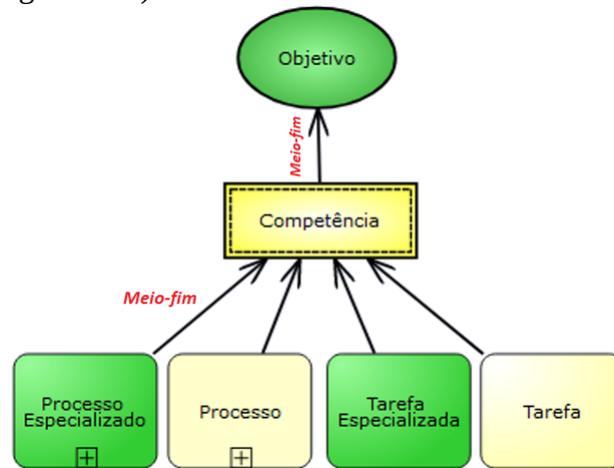


Figura 2.14. Exemplo conceitual do relacionamento de meios-fim (means-end).

A representação de múltiplos relacionamentos meios-fim expressa a *variabilidade* de operacionalizações para determinada competência, ou seja, existem diferentes formas de se alcançar a entrega de uma mesma competência. A partir do conceito de variabilidade é que se avaliam os melhores caminhos a se seguir, considerando as diferentes características das variações como contribuição de qualidades e adaptação ao contexto.

Na figura 2.15 apresentamos três exemplos de variabilidade de operacionalizações.

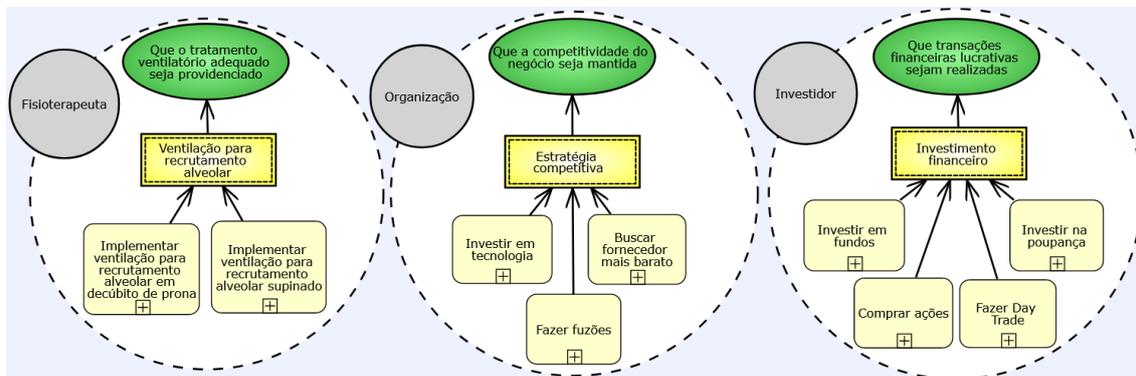


Figura 2.15. Exemplo de variabilidade de operacionalizações.

No primeiro exemplo, o Fisioterapeuta possui o objetivo “Que o tratamento ventilatório adequado seja providenciado”. Para satisfazer esse objetivo foi definida a competência “Ventilação para recrutamento alveolar”, a qual pode ser entregue a partir da implementação de duas operacionalizações (no caso, dois processos) que são: “Implementar ventilação par recrutamento alveolar em decúbito de prona” e “Implementar ventilação para recrutamento alveolar supinado”.

No segundo exemplo, a Organização possui o objetivo “Que a competitividade do negócio seja mantida”. Para satisfazer esse objetivo é necessário que a competência estratégia competitiva seja entregue. Três possíveis operacionalizações para essa

competência estão previstas: “Investir em tecnologia”, “Buscar fornecedor mais barato” e “Fazer fusões”.

No terceiro exemplo, o Investidor possui o objetivo “Que transações financeiras lucrativas sejam realizadas”. Para satisfazer esse objetivo é necessário que a competência “Investimento financeiro” seja entregue. As operacionalizações possíveis são: “Investir em fundos”, “Comprar ações”, “Fazer *Day Trade*” e “Investir na Poupança”.

As operacionalizações descritas apresentam diferentes formas de se implementar a competência. Cada operacionalização irá demandar um conjunto de requisitos próprios que podem diferenciar bastante o esforço de implementação, bem como o resultado.

Para se identificar qual a melhor operacionalização possível, há de se observar quais são as circunstâncias nas quais elas são melhor aplicáveis. Isso é o elemento central para a projeção de estratégias: identificar os possíveis estados de variáveis do ambiente para caracterizar situações para as quais determinadas operacionalizações são mais interessantes de se aplicar.

Em nosso exemplo, apresentamos primeiramente os conceitos presentes na definição da variabilidade (Figura 2.14 e Figura 2.15), entretanto, para a definição da estratégia, a avaliação de cenários e definição/adaptação de ações ocorrem da forma que for mais conveniente para a organização. As operacionalizações estratégicas devem ser intencionalmente definidas para lidar com situações do mundo real, as quais se alteram ao longo do tempo, demandando adaptações.

Na Administração, é comum a separação do ambiente interno e do ambiente externo para o estudo de possíveis cenários. O ambiente externo se torna mais relevante de se observar porque é onde ocorrem mudanças, as quais a organização não possui nenhum controle e que, possivelmente, irão requerer alguma adaptação. O ambiente interno abrange todo domínio da organização, em especial, os elementos que compõem a sua estrutura.

No entanto, é importante identificar o conjunto de elementos que se relacionam ao domínio que está sendo tratado, e que são capazes de influenciar de alguma forma a estratégia. O mapa desses elementos (*elementos do contexto*) delinea o *contexto* da estratégia, o que inclui o ambiente externo e interno.

As situações possíveis dentro de um contexto (*situação de contexto*) devem ser tratadas de forma estratégica através de operacionalizações

Na figura 2.16, foram incluídas as situações de contexto às respectivas operacionalizações.

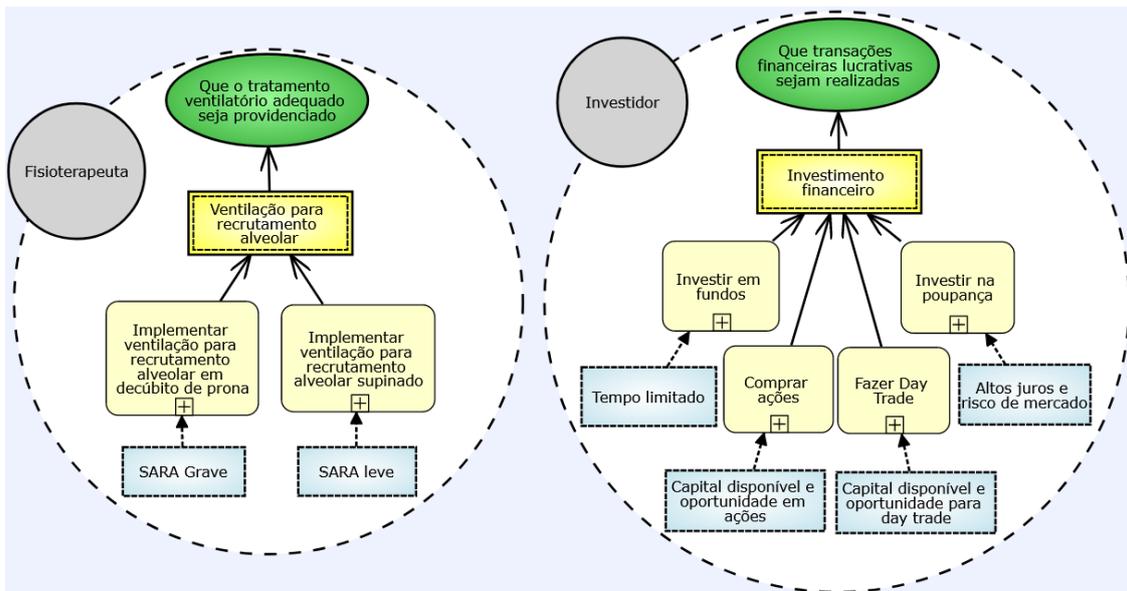


Figura 2.16. Exemplo de situações de contexto vinculadas a operacionalizações.

As situações de contexto “sugerem” a implementação das operacionalizações as quais estão vinculadas porque entende-se que estas operacionalizações foram planejadas de forma a conduzir ao melhor desempenho possível naquela situação.

2.4.4. Modelando as Situações de Contexto

O contexto é composto por um conjunto de elementos que estão direta e/ou indiretamente envolvidos em determinado domínio. O que define o contexto é, em última análise, o universo do discurso [Leite *et al.*, 2007].

Os elementos de um contexto podem ser neutros, ou capazes de interferir positivamente ou negativamente para outros elementos ou questões do domínio. Uma forma de facilitar a análise de contexto, é apenas considerar os elementos que podem causar interferência no domínio delimitado. A partir do mapeamento desses elementos, da avaliação de seus possíveis estados e respectivas contribuições, podem ser avaliadas possíveis situações no tempo. Isso permite a previsão de ações tendo em vista o melhor desempenho e/ou a redução de perdas.

O contexto é uma contribuição importante na tomada de decisões porque pode indicar momentos favoráveis ou desfavoráveis para realizar determinadas ações, ajudando na orientação organizacional. As situações de um contexto podem expressar o estado atual das coisas ou ser hipotéticas, com base em condições que, quando reais, indicam a vantagem ou a desvantagem em caminhos estratégicos. Por exemplo, a avaliação do contexto pode ajudar a decidir sobre investimentos e o redirecionamento de esforços para seguir melhores estratégias.

Na modelagem de processos (nível operacional), as situações de contexto podem ser caracterizadas pelos diferentes estados alcançados, os quais orientam o processo em diferentes direções [Santos *et al.*, 2011]. Entretanto, no âmbito estratégico, inúmeras variáveis internas ou externas à organização podem ser importantes, sendo essencial mapeá-las em um modelo específico que auxilie a expressar os elementos, seus estados e relações.

Considerando a complexidade no mapeamento dos elementos de contexto e na compreensão de como eles interagem com outros elementos, a linguagem GPI-HR utiliza um padrão para modelagem de contexto, considerando elementos já presentes na

linguagem (participantes do domínio que é expreso), seus estados e classificação de acordo com expectativas em uma avaliação de contexto.

Os componentes da linguagem GPI-HR definem os conceitos do domínio que podem ser modelados. Esses elementos delimitam o escopo dos conceitos do mundo real. Se o padrão for aplicado em outras linguagens, esses elementos devem ser revistos para os conceitos abordados na determinada linguagem.

Considerando a linguagem GPI-HR, identificamos os seguintes elementos que podem ser mapeados no contexto: a) indicadores (de qualquer natureza) que podem representar os valores das diferentes características dos elementos no contexto, b) o estado dos negócios, representando estados gerais não mensuráveis por indicadores, por exemplo, estados alcançados por uma instância de processo, c) estado dos recursos, representando a disponibilidade ou não de qualquer tipo de recursos, d) qualidade, representando a satisfação ou não de qualidades, que podem expressar pontos fortes e fracos, e) metas, representando metas cumpridas ou não e f) situações de contexto, representando sub contextos ou contextos relacionados que devem ser prioritariamente verdadeiros ou falsos.

Os recursos humanos também são variáveis do contexto e poderiam ser representados no modelo de contexto, entretanto, o GPI-HR apresenta recursos específicos para sua avaliação. Portanto, não os incluímos expressamente no modelo.

O modelo de contexto é genérico, apresentando o total de variáveis mapeadas ao contexto estratégico. Uma situação de contexto é uma *instância* deste modelo, normalmente caracterizada por um número limitado de elementos do ambiente que estão relacionados a uma estratégia e seus respectivos valores.

Portanto, o contexto é formado por um conjunto de elementos:

- *Contexto {Elemento A, Elemento B, Elemento C...}*

Sendo que cada elemento pode tomar um conjunto de valores no tempo:

- *Elemento A {valor1, valor2, valor3...}*

E a situação de contexto é caracterizada por um conjunto de elementos e seus respectivos valores caracterizados na determinada situação:

- *Situação do contexto {(ElementoA, valor2), (ElementoB, valor1), ...}*

Num **contexto** podem se estabelecer inúmeras **situações de contexto**.

O padrão de modelagem de contexto permite a classificação dos elementos do contexto. Essa classificação permite ampliar a possibilidade de registro de variáveis que estão vinculadas a situação de contexto de uma forma qualitativa positiva ou negativa, não sendo responsáveis por caracterizar o contexto, mas podem influenciar de alguma forma.

Registrar elementos de influência pode auxiliar a análise estratégica e a avaliação de situações do contexto, já que a influência destes elementos pode ser diversa, o que amplia o escopo de análise. Podem representar expectativas além do que se caracteriza como contexto favorável, ou apresentar pontos de atenção.

Os elementos do contexto podem ser classificados de acordo com quatro estados: mandatório ser verdadeiro, mandatório seja falso, desejável ser verdadeiro, desejável ser falso. Os elementos "mandatório ser verdadeiro" são necessários no contexto esperado, enquanto os elementos "obrigatórios sejam falsos" são proibidos. Os

elementos "desejáveis ser verdadeiro" contribuem positivamente para o contexto desejado, enquanto os elementos "desejáveis ser falso" contribuem negativamente.

A figura 2.17 apresenta o padrão de modelagem de contexto.

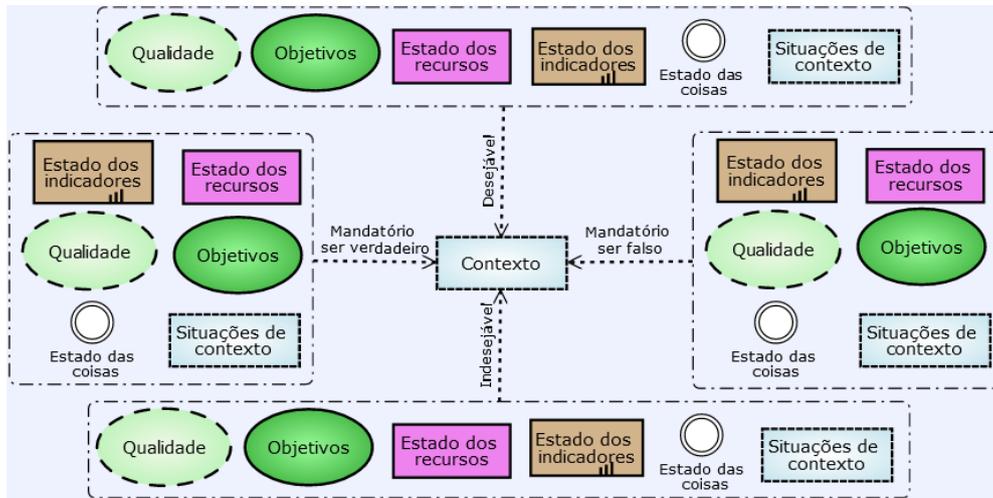


Figura 2.17. Padrão de modelagem de contexto.

2.4.5. Exemplo - Situações de Contexto

A Figura 2.18 apresenta um exemplo de modelagem da situação de contexto SARA grave, inicialmente apresentada na Figura 2.16.

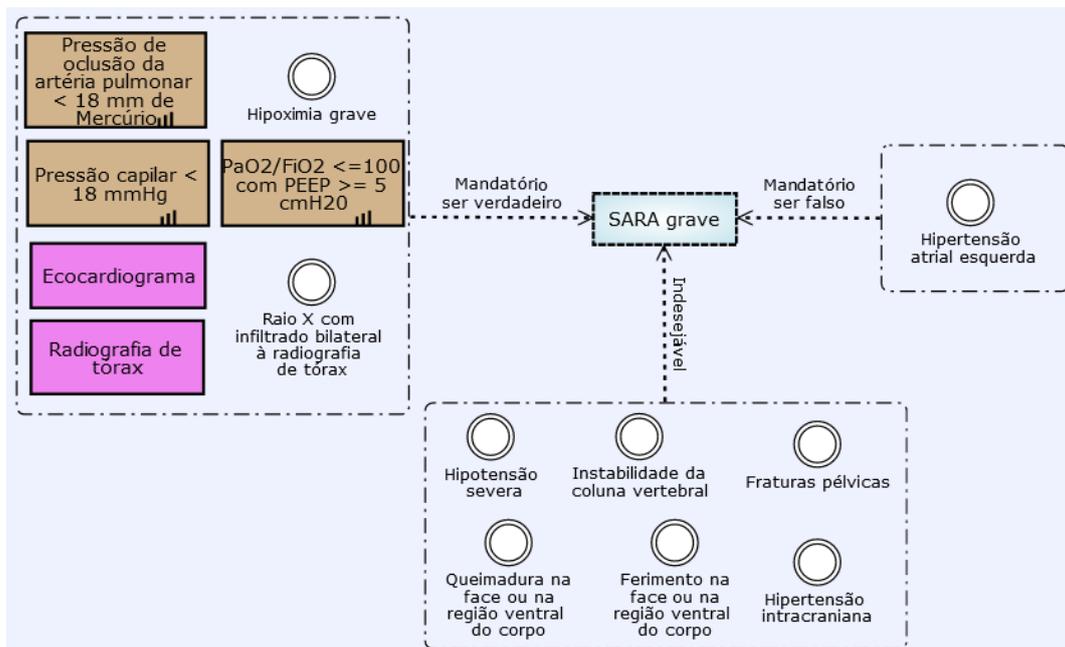


Figura 2.18. Situação de Contexto “SARA grave”, sem marcadores.

Particularmente, este contexto é composto pelos elementos obrigatórios (mandatórios) e um conjunto de estados indesejáveis. Neste caso, os elementos obrigatórios caracterizam o estado avançado da SARA (Síndrome da Angústia Respiratória Aguda). Os elementos indesejáveis não estão relacionados ao estado da doença, porém são indesejáveis pois irão dificultar o tratamento adequado, eventualmente demandando outros cuidados paralelos.

Este modelo representa apenas o mapa conceitual da situação de contexto SARA grave. Para se observar sobre o estado desta situação de contexto, é necessário

identificar os estados de todas as variáveis mapeadas. Para isso, são aplicados os marcadores no modelo.

2.4.6. O Uso de Marcadores nos Elementos do Contexto

Os elementos do contexto podem apresentar diferentes valores ao longo do tempo. Entretanto, para caracterizar uma situação do contexto, o intervalo dos elementos do contexto deve ser estabelecido. Após estabelecer os valores que caracterizam determinado elemento do contexto sob determinada situação, a sua avaliação torna-se binária, sendo verdadeira se seu valor estiver no intervalo ou falso, se estiver fora do intervalo.

Alguns elementos podem não ser caracterizados por um intervalo de valor, porém possuem estados bem definidos. Por exemplo, um objetivo, pode estar satisfeito ou não. Se satisfeito, ele denota uma variável do contexto em que seu estado é verdadeiro. Caso contrário, o seu estado é falso.

No caso dos recursos necessários para determinada situação de contexto, avalia-se a sua presença, o que caracteriza um estado verdadeiro, ou ausência, caracterizando o estado falso.

Os eventos/estados já possuem a sua característica binária. As situações de contexto, ou são verdadeiras ou falsas.

Os elementos do contexto são classificados conforme mostra a Tabela 2.8.

Tabela 2.8. Marcadores para elementos do contexto.

Marca	Semântica
	O estado do elemento é verdadeiro.
	O estado do elemento é falso.

2.4.7. O Uso de Marcadores na Situação de Contexto

Após definidos os marcadores dos elementos do contexto, é possível definir o marcador final da situação de contexto de acordo com o descrito na Tabela 2.7:

A Figura 2.19 apresenta um exemplo em que a SARA grave é uma situação real no contexto. Todos os elementos “Mandatário ser verdadeiro” estão marcados como verdadeiro (verde); todos os elementos “Mandatário ser falso” estão marcados como falso (vermelho); nos elementos indesejáveis, o estado de Fraturas Pélvicas é verdadeiro, indicando que há um elemento indesejável que está presente no contexto.

A presença deste elemento indesejável, define o marcador amarelo na situação de contexto, denotando que “O contexto é favorável à ação, mas deve ser verificado em termos de riscos potenciais de elementos indesejáveis com potencial contribuição negativa”.

A implicação estratégica da situação de contexto vai depender de outras variáveis presentes no modelo principal (Figura 2.16), onde o marcador amarelo irá propagar.

Tabela 2.7. Marcas para situação do contexto.

Marca	Semântica
	O contexto tem: Todos os elementos "obrigatório ser verdadeiro", verdade; Todos os elementos "desejáveis", verdade; Todos os elementos "indesejáveis", falsos; Todos os elementos "obrigatório ser falso", falsos. O contexto é totalmente favorável à ação.
	O contexto tem: Todos os elementos "obrigatório ser verdadeiro", verdade; Nenhum ou alguns elementos "desejáveis", verdade; Todos os elementos "indesejáveis", falsos; Todos os elementos "obrigatório ser falso", falsos. O contexto é favorável à ação.
	O contexto tem: Todos os elementos "obrigatório ser verdadeiro", verdade; Nenhum ou alguns elementos "desejáveis", verdade; Todos ou alguns elementos "indesejáveis", verdade; Todos os elementos "obrigatório ser falso", falsos. O contexto é favorável à ação, mas deve ser verificado em termos de riscos potenciais de elementos indesejáveis com potencial contribuição negativa.
	O contexto tem: Pelo menos um elemento "obrigatório ser verdadeiro", falso; Pelo menos um elemento "obrigatório ser falso", é verdade. Os elementos "desejáveis" e "indesejáveis" tornaram-se irrelevantes. O contexto é totalmente desfavorável à ação.

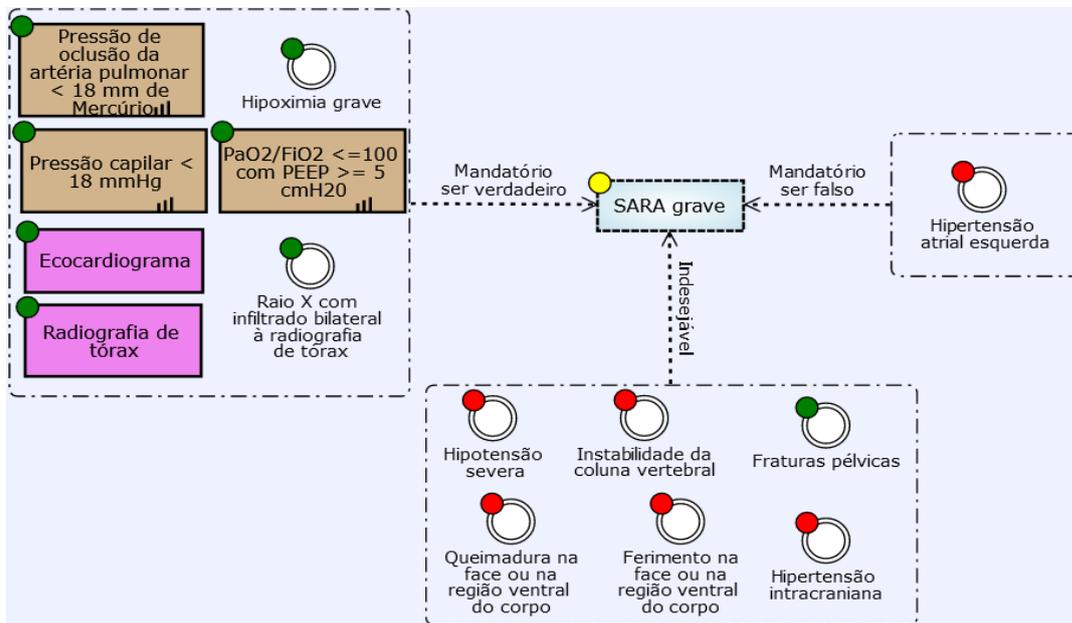


Figura 2.19. Situação de Contexto “SARA grave”, com marcadores.

2.4.8. Propagação de Marcadores ao Modelo Principal

Os marcadores dos modelos de Perfil de Competência e Situação de Contexto são propagados para o modelo principal, onde está descrita a variabilidade de operacionalizações para o determinado objetivo (Figura 2.20).

As propagações de alinhamento de RH têm sua origem nos Perfis Reais, através da alocação de um Recurso Humano em um papel da organização. Esse marcador

propaga para as tarefas e processos, prosseguindo para as respectivas competências e objetivos.

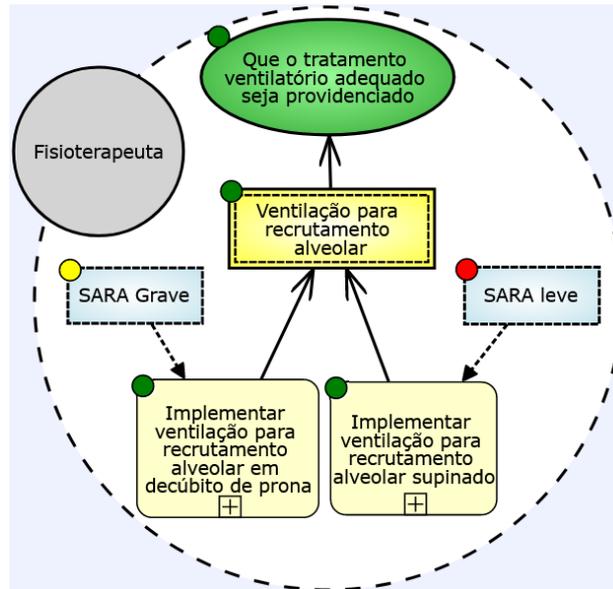


Figura 2.20. Propagação de marcadores até o objetivo.

Conforme já mencionado, o padrão de propagação do GPI-HR prioriza a propagação do marcador de menor valor. Essa estratégia de propagação auxilia a evidenciar problemas uma vez que eles sempre serão propagados em contrapartida aos marcadores superiores. Esta lógica de propagação também possibilita uma estratégia defensiva ao tomar decisões, já que sempre o pior caso será evidenciado. Nas análises de melhor caso, os marcadores de melhor posição devem ser desdobrados.

No exemplo da Figura 2.20, os processos estão com marcadores verdes, provenientes de um Perfil de Competências de um suposto RH. A marcação de Alinhamento de RH percorre o modelo propagando aos elementos superiores porque há nela uma conotação de impacto que pode ser positivo ou negativo.

Os marcadores da situação de contexto, assim como os marcadores de Alinhamento de RH, são elementos variáveis que servem como insumo para análises e decisões estratégicas. Os marcadores de situação de contexto não propagam, mas indicam se a operacionalização a qual ela está relacionada é plausível ou não de se implementar.

Neste caso, observa-se que a situação é de SARA grave, e que existem elementos relevantes a serem observados caso seja decidido implementar o processo “Implementar ventilação para recrutamento alveolar em decúbito de prona”. O elemento pode ser observado ao entrar no respectivo modelo de situação de contexto (Figura 2.19).

A variável de Alinhamento de RH, neste caso, é neutra, porque ambas as operacionalizações possuem a mesma marcação verde. Entretanto, se houvesse distinção entre as marcações, o processo decisório deveria considerar este marcador, ainda que o contexto fosse favorável, uma vez que o risco de impactos de desalinhamento de RH pode ser grande.

Na avaliação e decisão estratégica, os pormenores podem ser decisivos. Toda informação presente pode ter um papel importante se ajudar a evitar possíveis problemas futuros.

2.5. Modelagem de um Exemplo

Para consolidar o conteúdo apresentado até o momento, apresentamos uma descrição de caso, demarcada por tópicos, para facilitar o entendimento e a modelagem utilizando a linguagem GPI-HR.

O exemplo parte da perspectiva do ator Analista de empréstimos. Ele deve seguir uma estratégia definida para duas situações de mercado. O RH João é o indivíduo fictício definido neste exemplo. Seu perfil textual é apresentado para a interpretação do seu Perfil Real.

O exercício é seguido de uma proposta de modelagem para exemplificar o uso da linguagem. Seguem as descrições:

2.5.1. Estratégia

Descrição: *João é Analista de empréstimo. Seu principal objetivo é que suas transações sejam lucrativas. Sua transação principal é o empréstimo. Para alcançar sua meta, João implementa uma estratégia de concessão agressiva de empréstimos quando o mercado está aquecido. Entretanto, quando o mercado está em recessão, João aplica regras específicas para dificultar o crédito a clientes que possuem maior risco de inadimplência.*

Com essa descrição, foi possível modelar a Figura 2.21, contendo o objetivo, a competência que deve ser entregue e as duas possíveis formas de operacionalização, vinculadas a respectiva situação de contexto, definindo a sua estratégia em alto nível.

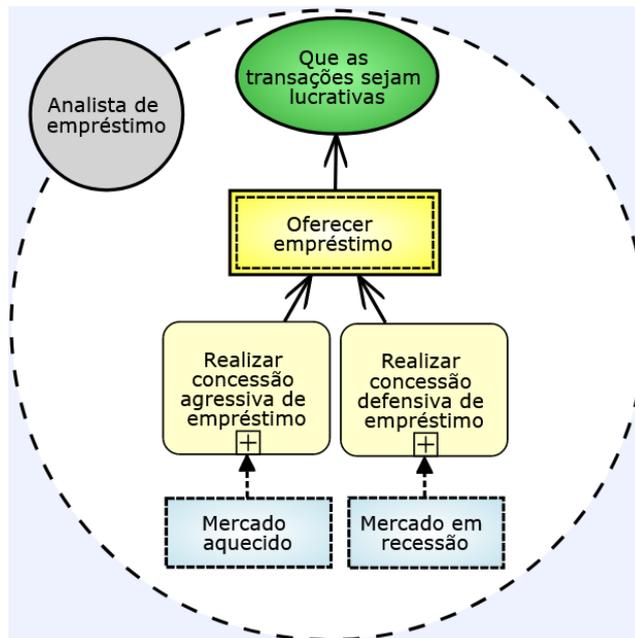


Figura 2.21. Modelo parcial apresentando as operacionalizações para “Oferecer empréstimo”.

2.5.2. Situação de contexto “Mercado aquecido”

Descrição: *O mercado está aquecido quando há baixa taxa de endividamento, baixa taxa de risco em empréstimos, baixa taxa de inadimplência no mercado, baixa taxa de juros e alta demanda por empréstimos. Também é necessário que a organização possua capital disponível para empréstimos e que não haja previsões de declínio no mercado. Às vezes existem políticas de subsídio governamental para empréstimos bancários, o que é bem-vindo. Apesar do ambiente interno estar adequado, há de se observar as*

tendências internacionais, uma vez que não é interessante que este mercado esteja em crise, considerando a possibilidade de impacto nos mercados. Também é relevante observar se a concorrência está muito agressiva, obtendo a clientela através de promoções que vinculem muito risco às transações.

Atualmente os indicadores de taxa de endividamento, taxa de risco em empréstimos, taxa de inadimplência no mercado, taxa de juros estão altas, e a demanda por empréstimos está baixa. Externamente, há precisões de declínio do mercado, apoiado por um ambiente internacional em crise, o que é indesejado. Até o momento isso não afetou a nossa fatia de mercado em valor relativo, já que todas as empresas do ramo estão reduzindo sua carteira de clientes em valor absoluto. Isso não é diferente em nossa empresa, por isso precisamos observar. Infelizmente o governo cortou os subsídios provenientes do FGTS e da poupança, mas possuímos capital para empréstimos reservado.

Através da descrição foi definido o modelo de situação de contexto “Mercado aquecido”, conforme mostra a figura 2.22.

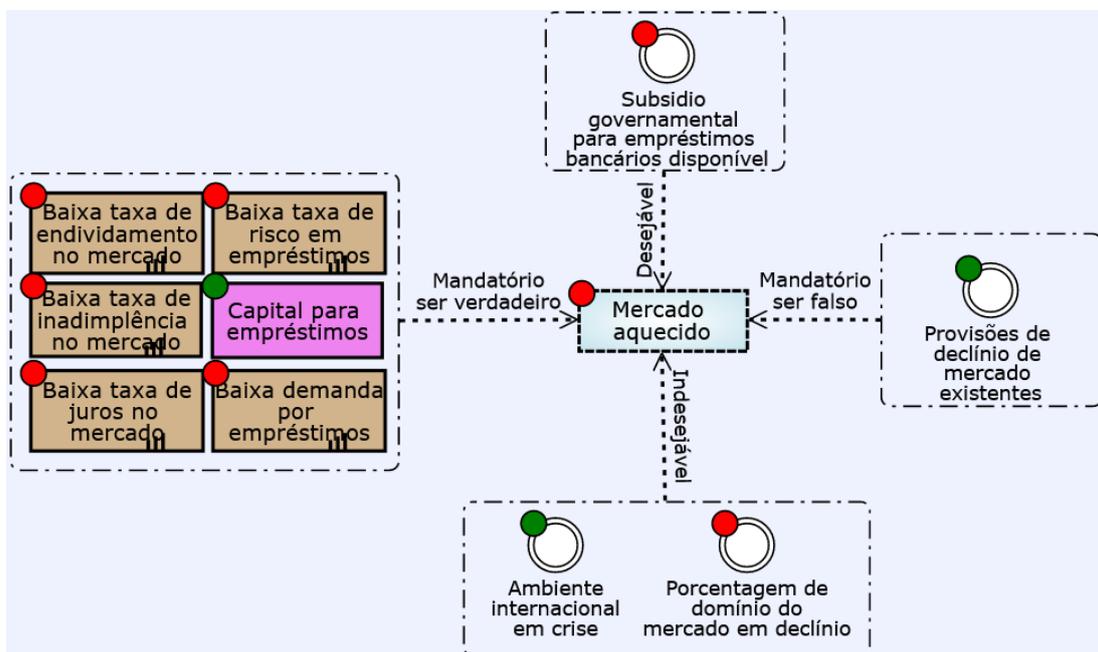


Figura 2.22. Situação de Contexto “Mercado Aquecido”.

2.5.3. Situação de Contexto “Mercado em Recessão”

Descrição: O mercado está em recessão quando há alta taxa de endividamento, alta taxa de risco em empréstimos, alta taxa de inadimplência no mercado e alta taxa de juros. Para emprestar nesse cenário, a margem anual de perda de lucro seja menor que 5%. Também é necessário que a organização possua capital disponível para empréstimos. Especificamente durante a recessão, é interessante se houver alta demanda a partir de clientes “especiais”. Se houver competidores atuando de maneira agressiva, é importante observar os motivos. Eventualmente há alguma oportunidade no mercado que nossos analistas não identificaram. Não é interessante termos muitos clientes comuns pedindo empréstimo, porque todos serão negados, potencialmente causando insatisfação. Isso deverá ser observado de perto.

Nossos clientes “especiais” são a melhor opção para fechar contrato de empréstimos, já que o risco de calote é bem mais baixo do que entre os clientes comuns. Infelizmente, esta demanda não está alta. Mas até entre clientes comuns a demanda está baixa, e isso tem feito com que os competidores atuem agressivamente, se expondo

ao risco. Trabalhamos para que a nossa perda de lucro seja menor do que 5% quando há movimentos negativos no mercado, mas este ano esse valor já foi superado especialmente pela baixa procura por empréstimos em todas as classes sociais, motivada pela alta taxa de juros.

Através da descrição foi definido o modelo de situação de contexto “Mercado aquecido”, conforme mostra a figura 2.23.

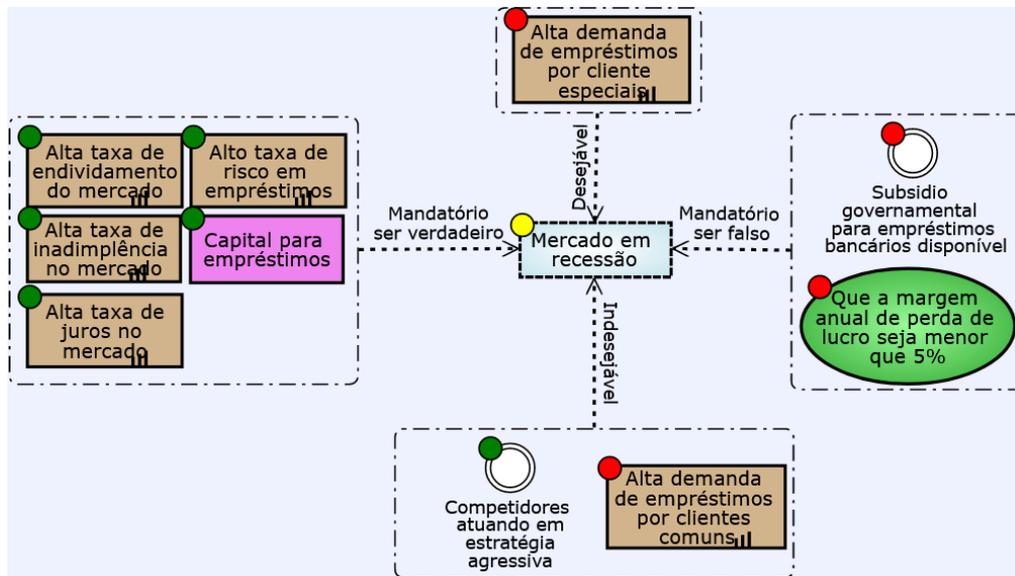


Figura 2.23. Situação de Contexto “Mercado em Recessão”.

De acordo com a descrição, o mercado encontra-se em recessão. Isso é definido pelas variáveis mandatórias, as quais se enquadram nesta situação de contexto. Entretanto, há um ponto de atenção porque um elemento indesejável se faz presente. Os competidores estão atuando de forma agressiva em um ambiente de baixa. Isso pode ser uma medida necessária para que a empresa sobreviva em tempos difíceis, no entanto, essa decisão implica em riscos. Por isso, o marcador resultante da situação de contexto é amarelo. Esse marcador indica duas coisas: a primeira é que a situação de contexto é verdadeira; a segunda é que há elementos indesejáveis presentes no cenário, implicando na necessidade de se observar com cuidado qual é esse elemento e o motivo dele estar presente. Essas informações podem ser determinantes para uma melhor decisão estratégica. Nenhum elemento desejável está presente, mas, ainda que estivesse presente, o marcador permaneceria amarelo, visando expor os elementos indesejáveis presentes.

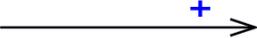
2.5.4. Requisitos do Processo Concessão defensiva de empréstimo

Descrição: Durante um período de recessão, quando João recebe uma demanda de empréstimo, ele segue o processo “Concessão defensiva de empréstimo”. Neste processo são implementadas algumas regras específicas para a concessão de empréstimo, visando reduzir riscos inerentes de um cenário de mercado em recessão. A primeira regra é que o empréstimo seja concedido apenas para os clientes especiais, uma vez que o grau de risco de não pagamento destes clientes é bem menor do que as dos outros clientes. Além disso, vinculamos o empréstimo a garantia de algum bem penhorável. Os clientes especiais, no geral, possuem pelo menos um patrimônio penhorável, mas assim mesmo devemos verificar antes de conceder empréstimos se o bem ainda está disponível, se possui valor compatível para cobrir o valor de empréstimo solicitado, e se está livre e desembaraçado. Outra regra importante é

garantir que o empréstimo aprovado seja depositado diretamente em uma conta nominal, para que possamos ter o comprovante de depósito. Alguns clientes informam contas de terceiros e pessoa jurídica, e o depósito nestas contas nos trouxe problemas no passado. Nestes períodos de recessão não temos produtos disponíveis para os clientes que não sejam especiais, então recusamos com os devidos cuidados e atenção ao pedido dos clientes comuns, visando mantê-los em nossa carteira de clientes.

A partir da descrição acima, foi definido o exemplo presente na Figura 2.24, estendendo a Figura 2.21.

Tabela 2.7. Elementos incluídos no modelo estratégico.

	Representa uma qualidade que pode ser satisfeita em diferentes níveis, dependendo da visão do avaliador.
	Representa um relacionamento de decomposição (semântica de “E” ou “AND”).
	Representa um relacionamento de contribuição positiva. Especificamente utilizado para relacionar operacionalizações a objetivos qualitativos.

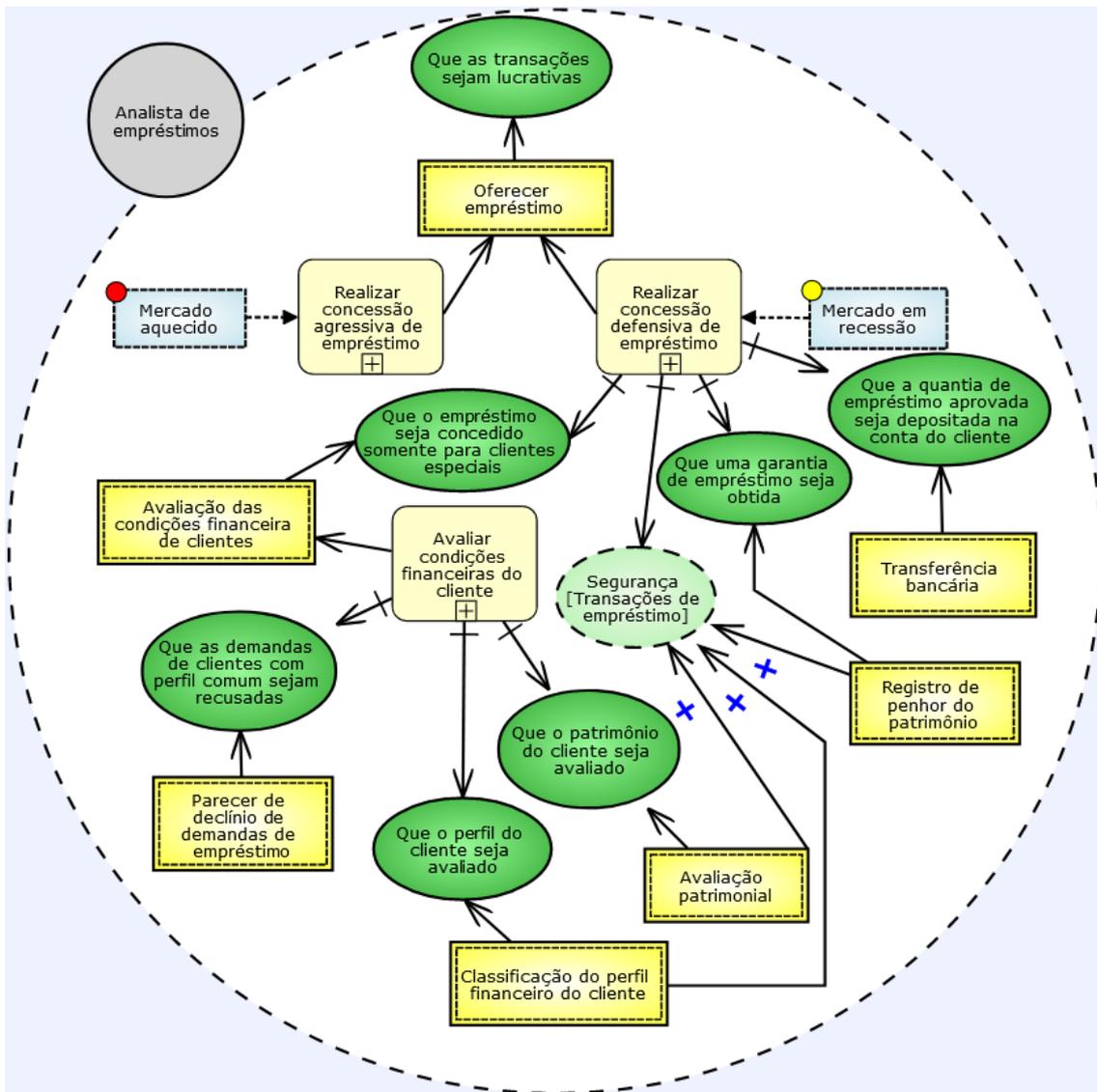


Figura 2.24. Modelo apresentando o detalhamento da operacionalização “Realizar concessão defensiva de empréstimo”.

O conjunto de regras existentes dentro do processo de realizar concessão defensiva de empréstimo são requisitos que devem ser considerados na projeção das tarefas do processo. Esses requisitos são registrados como objetivos que decompõem o processo “Realizar concessão defensiva de empréstimo”, os quais devem ser satisfeitos ao longo da sua execução, à medida em que são entregues competências. O mapeamento desses objetivos detalha o que deve ser alcançado no nível operacional. As competências que deverão ser entregues para satisfazer os objetivos foram projetadas e relacionadas a eles.

Durante a execução do processo, essas competências serão entregues, satisfazendo todos os objetivos mapeados. Portanto, as tarefas descritas ao longo do processo são referentes às operacionalizações destas competências.

O processo “Avaliar condições financeiras do cliente” é um subprocesso de “Realizar concessão defensiva de empréstimo”, compondo-o juntamente com outras tarefas. Este subprocesso foi decomposto para explicitar três objetivos que devem ser alcançados durante a avaliação das condições financeiras do cliente. Por sua vez, esses objetivos estão relacionados às competências que os operacionalizam. Neste nível, essas

competências não permitem mais decomposição, devendo ser relacionadas diretamente às partes do processo que a operacionalizam (será apresentado mais a frente).

2.5.5. Processo (operacionalização Concessão defensiva de empréstimo)

Descrição: Após receber a requisição, seu primeiro passo é classificar o perfil do cliente. Se o cliente for do tipo “Especial”, ele verifica a existência de patrimônio para penhor (uma forma de garantia do empréstimo). Se o cliente possuir patrimônio disponível, ele registra o patrimônio como garantia de empréstimo no cartório. Posteriormente, deposita o valor do empréstimo diretamente na conta do cliente. Se o cliente não for “especial” ou não possuir patrimônio para penhora, a demanda de empréstimo é declinada e um parecer de declínio de empréstimo é emitido ao cliente.

Partindo desta descrição, o processo apresentado na Figura 2.25 foi construído:

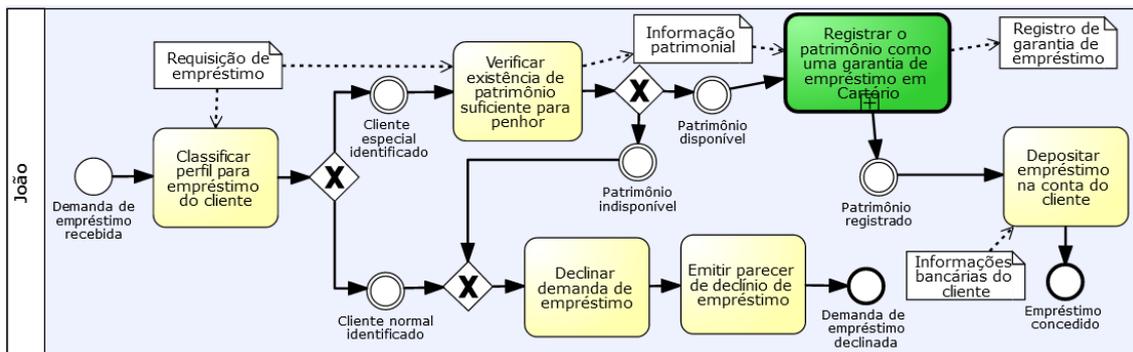


Figura 2.25. Modelo apresentando o detalhamento da operacionalização “Realizar concessão defensiva de empréstimo”.

2.5.6. Perfil de Competências para o ator Analisa de Empréstimos

Descrição: Para atuar como analista de empréstimos, é necessário que o profissional tenha conhecimentos das regras internas de análise e classificação de perfis de cliente, bem como os valores patrimoniais do mercado atualizados. Não é fácil avaliar um patrimônio de forma adequada. O profissional deve ter um olhar minucioso e técnico. Nos processos internos, ele precisa saber trabalhar no computador, fazer consultas online, especialmente porque todos os nossos softwares são vinculados ao sistema bancário. É importante que o profissional consiga medir os riscos de um cliente em potencial para não termos problemas com inadimplência na carteira de empréstimos. Nós avaliamos o desempenho do profissional através de um indicador de inadimplência.

Especificamente nos processos em que existam registros patrimoniais no cartório, o profissional necessitará ser meticuloso e responsável, porque há prazos, os documentos devem estar impecáveis e, em caso de problema, ele deve estar apto a resolvê-los da melhor forma possível. Em cartórios, problemas são comuns, mas para tudo há um jeito.

A partir desta descrição, foram modelados dois Perfis Requeridos, sendo um referente às tarefas gerais, e outro referente a tarefa especializada “Registrar o patrimônio como uma garantia de empréstimo em cartório”, a qual foi modelada como uma tarefa especializada (verde) no processo descrito na Figura 2.25.

Em ambos os perfis todos os elementos CHA foram modelados como obrigatórios. No perfil da tarefa especializada (Figura 2.26), foram definidos dois indicadores relacionados com a aptidão de responsabilidade para avaliar o desempenho

do profissional no trato com os procedimentos e documentos cartoriais, uma vez que se trata de variáveis importantes para o bom desempenho da tarefa.

No perfil das tarefas básicas (Figura 2.27), foi definido um indicador para medir o desempenho da inadimplência na carteira de clientes como uma forma de avaliar a aptidão de mensuração de risco.

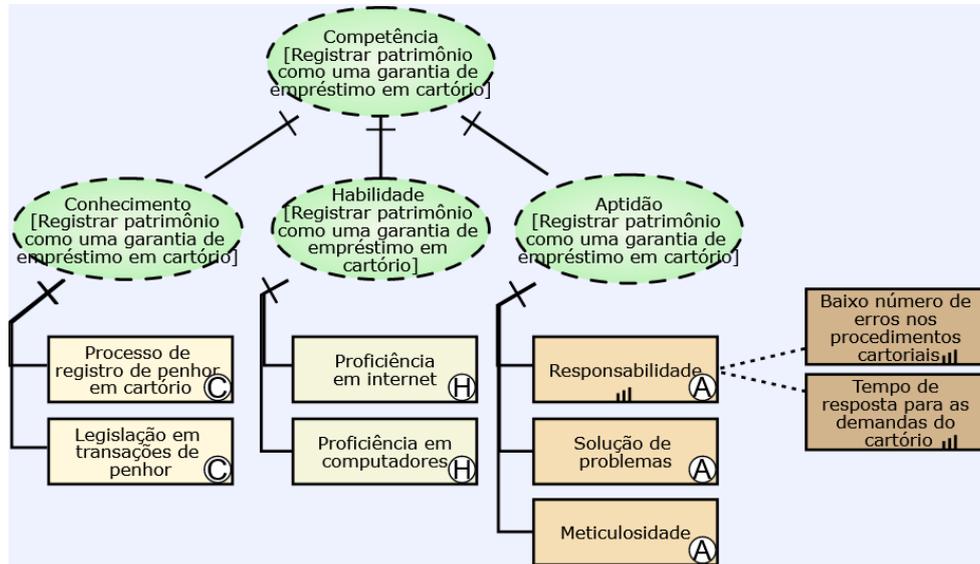


Figura 2.26. Perfil Requerido da tarefa “Registrar patrimônio como uma garantia de empréstimo em cartório”.

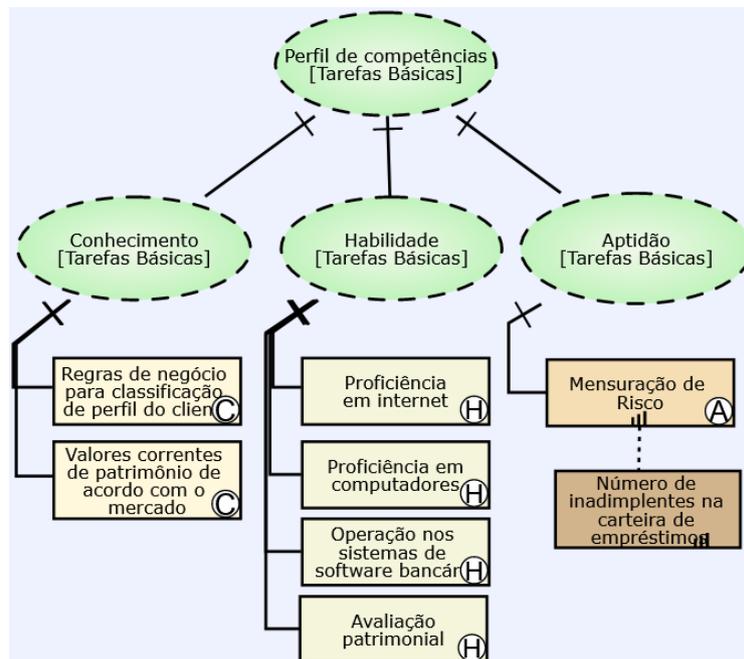


Figura 2.27. Perfil Requerido das tarefas básicas no processo “Realizar concessão defensiva de empréstimo”.

2.5.7. Perfil Profissional de João para Definição do Perfil Real

Descrição: João é um excelente profissional, com grandes conhecimentos e habilidades. Não existem inadimplentes em sua carteira, e nos procedimentos cartoriais há baixo número de erros, entretanto, João anda um pouco relapso, perdendo prazos no cartório. Isso tem atrasado demasiadamente pedidos de empréstimo, fazendo com que o

cliente desista. Além disso, perdemos dinheiro também para reabrir processos e refazer procedimentos.

Partindo da descrição do perfil profissional de João, foram construídos o Perfil Real das tarefas básicas (2.28) e o Perfil Real da atividade especializada (2.29). Em sua descrição, é possível observar que nos quesitos de conhecimentos e habilidades ele se encontra favorável, portanto, todas as marcações foram definidas como verde. Na aptidão de mensuração de risco, observa-se que João possui uma carteira impecável, sem nenhum registro de inadimplência, o que justifica o marcador azul em seu indicador de “Número de inadimplentes na carteira de empréstimo”. Por consequência, esse marcador propaga para a aptidão “Mensuração de risco”. O marcador final de Alinhamento de RH de João neste perfil é verde, demonstrando que ele possui competência para lidar com as tarefas básicas, apresentando o desempenho esperado no seu perfil de conhecimento e habilidades, e um desempenho superior no perfil de aptidão.

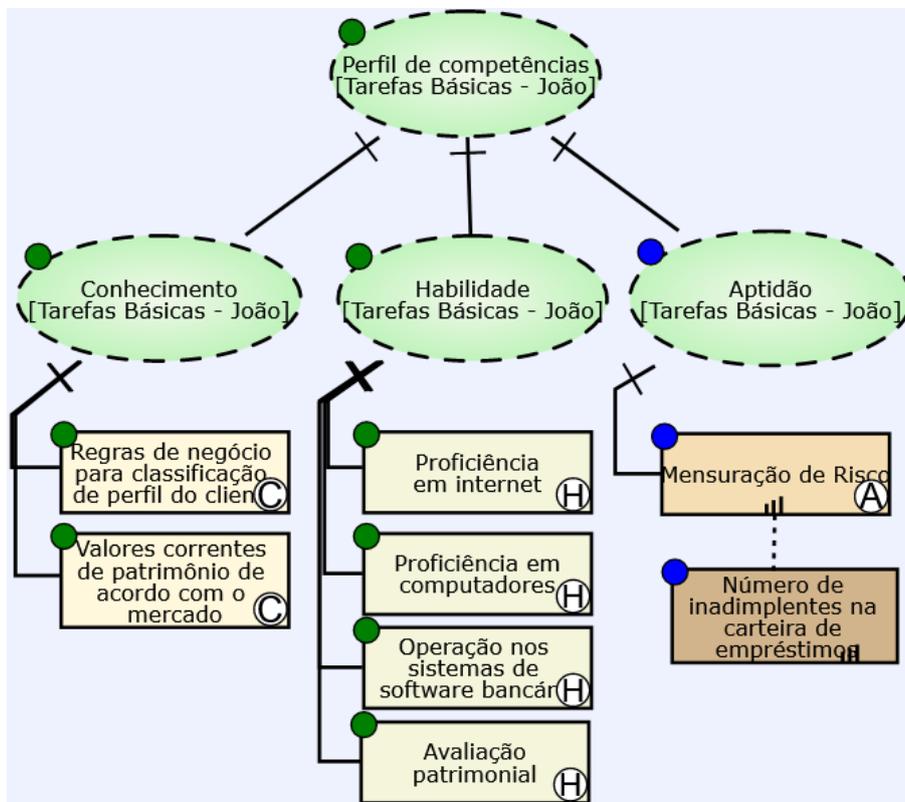


Figura 2.28. Perfil Real de João referente às tarefas básicas no processo “Realizar concessão defensiva de empréstimo”.

No Perfil Real da tarefa especializada, o Perfil de conhecimento e habilidade seguiu as boas anotações, conforme no perfil anterior. Entretanto, na aptidão de responsabilidade há a descrição de problemas provenientes da perda de prazos no cartório. Por isso foi definido o indicador “Tempo de resposta para as demandas do cartório” como um indicador de responsabilidade. Por ser um relato de problema, utilizamos o marcador amarelo para este indicador, denotando o desempenho abaixo do esperado. O marcador poderia ser interpretado como vermelho, mas pela descrição, não parece se tratar da ausência grave do quesito de responsabilidade. Somente um problema isolado que possui potencial de ser contornado.

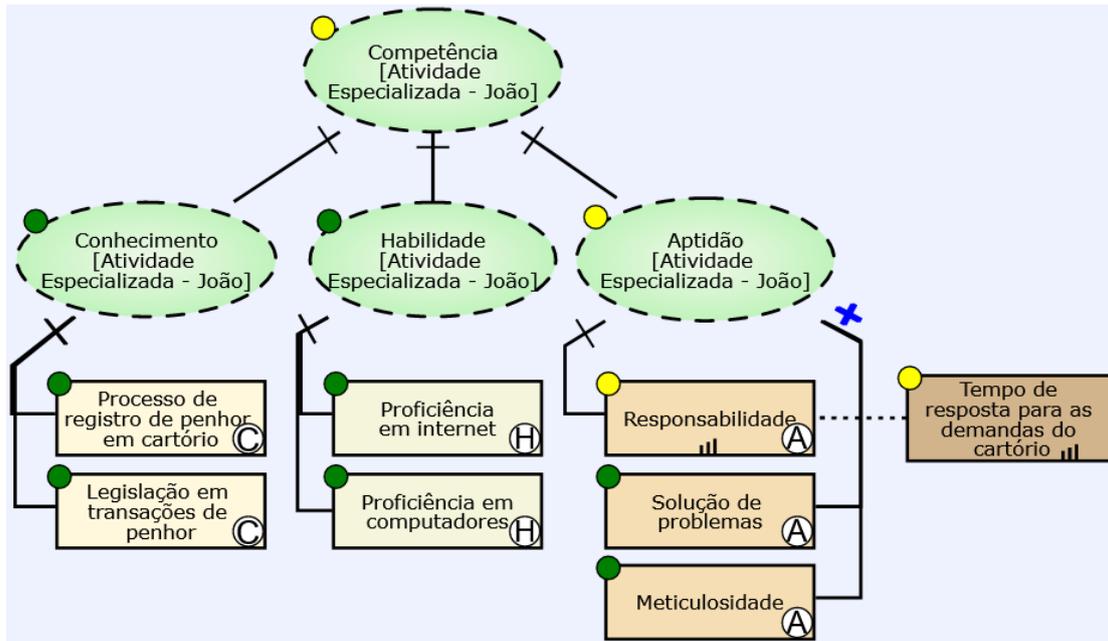


Figura 2.29. Perfil Real de João referente a atividade especializada no processo “Realizar concessão defensiva de empréstimo”.

Ambos os perfis podem ser consolidados em um único perfil (Figura 2.30). Entretanto, separá-los auxiliará na identificação de seus impactos nos outros modelos. O marcador final de Alinhamento de RH para o João pode ser visto na Figura 2.30, que apresenta o exemplo de Perfil Real unificado.

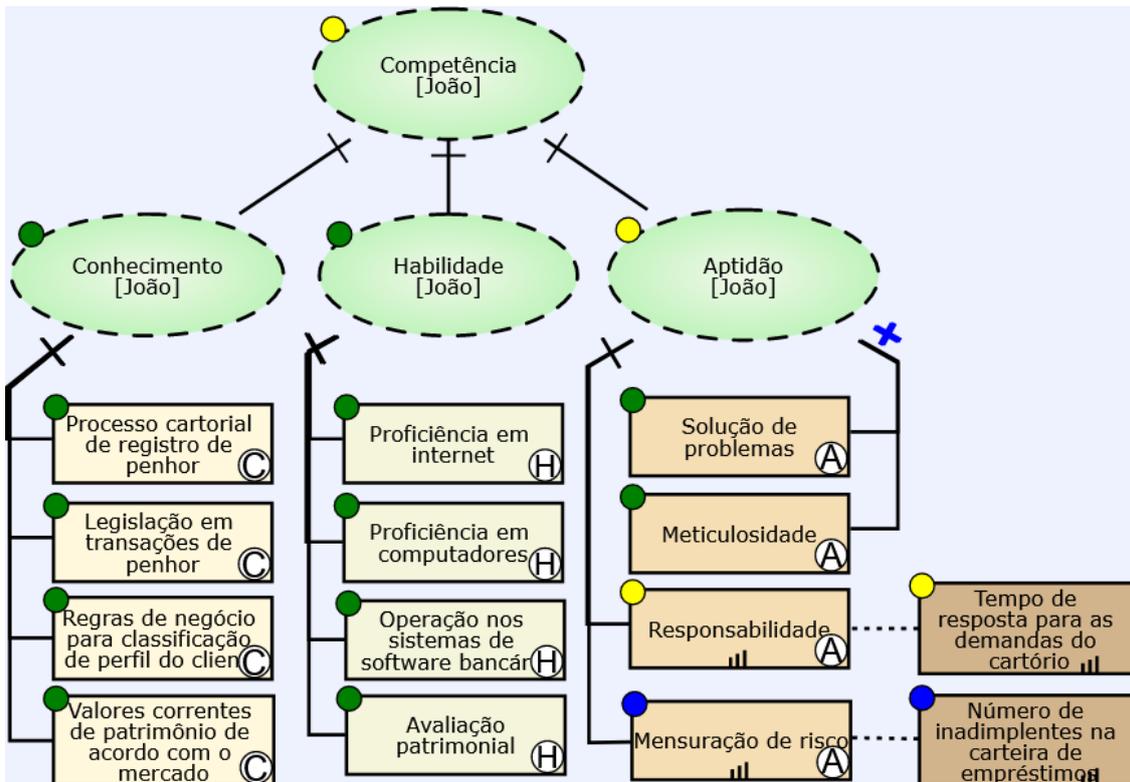


Figura 2.30. Perfil Real de João no papel de Analista de Empréstimos no processo “Realizar concessão defensiva de empréstimo”.

2.5.8. Alinhamento Entre Modelos e Propagação de Marcadores

A tarefa de alinhamento entre modelos consiste na definição de relacionamentos entre o modelo estratégico (Figura 2.24) e o modelo operacional (Figura 2.25). Neste momento é possível identificar possíveis desalinhamentos presentes nos modelos, o que pode ser causado por fatores diversos como, por exemplo, ausência de informação, entendimento inadequado e desentendimentos de fontes de informação (quando há mais de uma).

O alinhamento entre modelos visa a relação entre tarefas do processo e respectivas competências. O alinhamento ocorre quando toda tarefa estiver vinculada a uma competência, e toda competência estiver vinculada a pelo menos uma tarefa. Desta forma não podem existir tarefas e competências sem relação, ou teremos tarefas sem justificativas de existir, ou competências sem operacionalização no processo. Caso ocorra um dos casos, é necessário identificar o motivo, obter as informações adequadas e corrigir o modelo.

Durante o mapeamento destes relacionamentos, há a possibilidade de críticas dos modelos, portanto esse procedimento também é uma oportunidade de verificar a correteza e completeza do modelo, além de possibilitar identificar problemas de entendimento e qualidade das informações pela perspectiva do modelador e de suas fontes de informação (*stakeholders*).

Após o mapeamento dos relacionamentos, será possível utilizá-los como pontes para a propagação dos marcadores presentes no modelo de processos para o modelo estratégico. Esses marcadores irão representar potenciais impactos por (des)alinhamento de RH.

A Figura 2.31 apresenta o resultado do alinhamento e propagação dos marcadores. Utilizamos o recurso de “agrupamento” para relacionar duas tarefas a uma competência. A propagação seguiu a regra do pior caso, resultando um marcador amarelo no objetivo principal “Que as transações sejam lucrativas”.

Este modelo final é o *Diagrama Integrado* do GPI-HR [Sousa e Leite, 2017]. A partir dele são feitas as considerações estratégicas, tendo como insumo, além do detalhamento da operacionalização no modelo estratégico e no modelo operacional, as informações de Alinhamento de RH, Situação de Contexto, e as opções de operacionalização para a entrega da competência que satisfaz o objetivo principal.

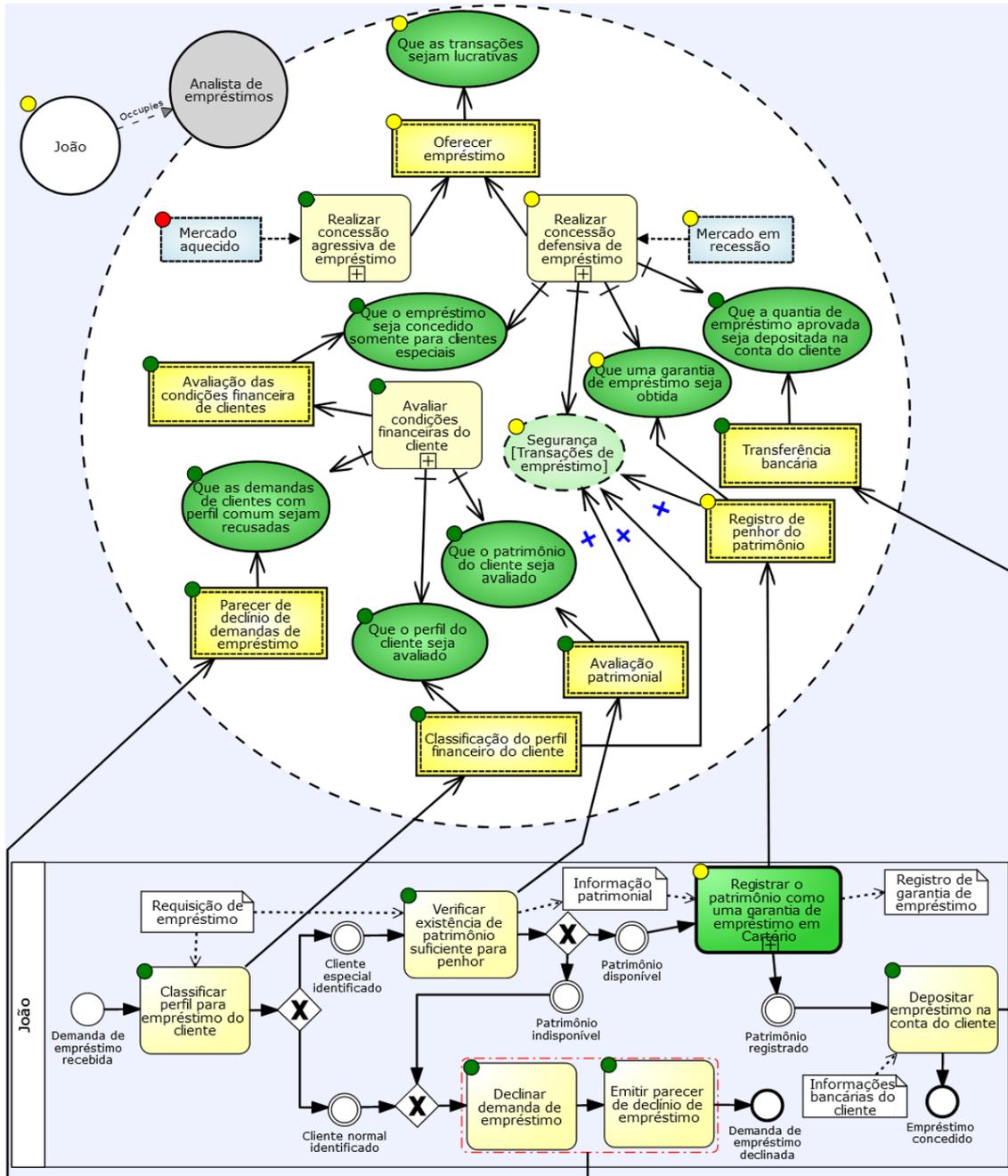


Figura 2.31. Diagrama Integrado.

2.5.9. Análise do Alinhamento de RH e Estratégia

A análise do alinhamento apresentada aqui é um exemplo. É possível obter diferentes interpretações e conclusões a partir do mesmo modelo.

O caso utilizado como exemplo demonstra um cenário de em que o mercado está em recessão. A indicação neste caso é a aplicação da operacionalização “Realizar concessão defensiva de empréstimo”. Adicionalmente, temos um marcador amarelo nesta operacionalização, e na situação de contexto. Portanto, há a necessidade de se avaliar as causas.

Ademais, observa-se que a operacionalização “Realizar concessão agressiva de empréstimo” está com o marcador de alinhamento de RH verde, demonstrando maior probabilidade de sucesso na implementar desta operacionalização do que na outra. Isso

também pode ser relevante na decisão estratégica, ainda que a situação de contexto indique que não está em mercado aquecido.

Observando através do modelo de situação de contexto, identifica-se que o há um elemento de contexto indicando que os competidores estão utilizando uma estratégia agressiva para obter clientes. Isso é indesejável em uma situação de mercado em recessão, pois neste momento a organização atua estrategicamente de forma defensiva, e o tom agressivo dos competidores pode piorar o enfrentamento da crise. Por outro lado, há de se verificar o motivo desse comportamento porque, talvez, os competidores possam ter observado algo relevante no mercado para atuar desta forma, o que é de interesse desta organização.

Quanto ao marcador de Alinhamento de RH, observa-se que há um desalinhamento pontual, por falta de desempenho em um indicador de responsabilidade. No entanto, isso pode afetar a concessão de empréstimo, porque faz referência a uma tarefa fundamental na concessão defensiva de empréstimo, e está relacionado à segurança nas transações e garantia do capital emprestado pela organização. A área de RH deve atuar neste desalinhamento, entretanto, há muitas possibilidades de intervenção. Por exemplo, atuar diretamente com o indivíduo para identificar o que pode estar motivando esse marcador.

É possível verificar que o perfil do indivíduo é bom, e há o desvio pontual no tempo de resposta para as demandas de cartório. Portanto, podemos entender que vale à pena investir no indivíduo João. Eventualmente seu problema é pessoal, e isso não é algo que se tornará explícito no dia a dia. A área de RH pode acompanhar problemas pessoais para apoiar o retorno deste recurso humano ao seu melhor desempenho. Para saber se essa seria uma boa solução, uma forma seria observar o histórico do recurso humano. Ele apresenta o problema de responsabilidade neste momento ou sempre apresentou? Essa informação não temos no modelo.

Outra ação é a chamada formal da melhoria deste aspecto, de forma transparente e profissional, fazendo com que o funcionário se conscientize do seu papel e impacto de eventuais erros para a organização.

Outras ações são possíveis, como a demissão ou substituição do recurso humano. Entendemos que um problema pontual pode ser resolvido com menos atrito.

Por outro lado, a implementação da estratégia agressiva também pode ser uma opção, uma vez que o João apresenta marcador verde nesta operacionalização, e a concorrência, por algum motivo, está aplicando esta estratégia. Se essa motivação for plausível para a organização, essa pode ser uma decisão satisfatória, desviando dos problemas gerados na concessão defensiva de empréstimo (pelo menos por enquanto), e se alinhando à concorrência. Como há uma baixa procura por empréstimos, de forma geral, talvez essa seja a motivação dos competidores para atuar agressivamente.

Para que a organização não perca negócios em um momento difícil, talvez seja necessário atuar agressivamente, mesmo em um cenário de recessão e alto risco. A organização possui capital disponível, o que é um dos aspectos mais importantes nos tempos de crise.

Portanto, indicamos a possibilidade de se atuar de forma agressiva, considerando o marcador de Alinhamento de RH, para desviar do possível impacto negativo na estratégia defensiva, ao menos de uma forma imediata, enquanto trata do alinhamento

do João. E a partir da interpretação das variáveis presentes nos modelos de situação de contexto, que mostram um cenário de baixa procura por empréstimos aliado às ações dos competidores em atuar de forma agressiva, ou seja, utilizando recursos para facilitar o acesso ao empréstimo visando obter clientes, ainda que existam riscos maiores envolvidos.

Baseado neste modelo, estas são as considerações que apresentamos como exemplo.

2.6. Conclusão

Este minicurso apresentou a Modelagem da Estratégia e Alinhamento de Recursos Humanos Baseado em Competências utilizando a linguagem GPI-HR. Foram utilizados casos simplificados para auxiliar no entendimento dos conceitos e uso da linguagem. A aplicação da linguagem em cenários reais pode demandar a atuação de profissionais de diversas áreas.

Para a construção de um modelo mais complexo, observamos que a linguagem envolve aspectos de diversas áreas, como a modelagem de processos e objetivos, a modelagem de perfis de competência, e a projeção e modelagem de estratégias. Esses modelos também podem envolver diversas áreas, sendo necessário elicitar informações a partir de diferentes grupos.

O trabalho de construção de um modelo amplo demanda minuciosa fase de elicitação para se obter informações suficientes para a construção adequada dos modelos. A qualidade dos modelos poderá influenciar nas interpretações que ele auxiliará a obter.

Portanto, a complexidade do modelo está diretamente associada à complexidade do domínio organizacional. Vinculada a análise estratégica, toda informação pode ser útil. Observamos que o Alinhamento de RH é apenas um aspecto na composição organizacional e que outras variáveis são muito importantes para a análise estratégica, como informações provenientes de outras áreas, por exemplo, de Finanças, Produção, Fornecimento, *Marketing*, Tecnologia da Informação, Jurídico, Comercial e Fiscal (não limitada a estes exemplos).

A linguagem GPI-HR se encontra em constante evolução, visando incluir de forma incremental, mais conceitos organizacionais pertinentes à análise estratégica e de alinhamento organizacional.

Estudos orientados a linguagens de domínios específicos são atualmente pesquisados visando extrair conhecimentos importantes para a definição de linguagens que apoiem a modelagem, facilite o entendimento e possibilite a análise através de modelos.

Referências

- AZEVEDO, C. L.; IACOB, M. E.; ALMEIDA, J. P. A.; VAN SINDEREN, M.; PIRES, L. F.; GUIZZARDI, G.; “Modeling resources and capabilities in enterprise architecture: a well-founded ontology-based proposal for ArchiMate”; *Information Systems*, 54, 235-262, 2015.
- BARNEY, JAY B. Integrating organizational behavior and strategy formulation research: A resource-based analysis. 1992.
- KLINK, Marcel van der; BOON, Jo. The investigation of competencies within professional domains. *Human resource development international*, v. 5, n. 4, p. 411-424, 2002.
- BRANDÃO, H.; PENA, G.; TOMÁS, DE A.; “Gestão de competências e gestão de desempenho: tecnologias distintas ou instrumentos de um mesmo constructo?”; *Revista de Administração de Empresas*, São Paulo, v.41, n.1, p.08-15, jan/mar, 2001.
- CARBONE, P.; BRANDÃO, H.P.; LEITE, J.B.D., VILHENA, R.M.P.; “Gestão por competências e gestão do conhecimento”; Rio de Janeiro: Ed. Fundação Getúlio Vargas, 2005.
- CURTIS, B.; HEFLEY, B.; MILLER, S.; “People Capability Maturity Model”; Software Engineering Institute; 2009.
- DRAGANIDIS, F.; MENTZAS, G.; “Competency based management: a review of systems and approaches”; *Information management & computer security*, 14(1), 51-64, 2006.
- DUTRA, Joel Souza . A gestão de pessoas articulada por Competências. *Revista T&D Inteligência Corporativa* , v. 15, p. 20-21, 2007.
- FLEURY, Maria Tereza Leme; FLEURY, Afonso. Em busca da competência. *Encontro de Estudos Organizacionais*, v. 1, 2000.
- FLEURY, Maria Tereza Leme; FLEURY, Afonso. Construindo o conceito de competência. *Revista de administração contemporânea*, v. 5, p. 183-196, 2001.
- FLEURY; M. T. L.; FLEURY; A. C. C.; “Alinhando estratégia e competências”; *RAE-Revista de Administração de Empresas*, vol. 44, n. 1, 2004.
- FLEURY, Afonso; FLEURY, Maria Tereza Leme. *Brazilian multinationals: competences for internationalization*. Cambridge University Press, 2011.
- GRANT, R. M.; “Toward a knowledge-based theory of the firm”; *Strategic management journal*, 17(S2), 109-122, 1996.
- LIASKOS, S., JIANG, L., LAPOUCHNIAN, A., WANG, Y., YU, Y., DO PRADO LEITE, J. C. S., & MYLOPOULOS, J. (2007). Exploring the Dimensions of Variability: a Requirements Engineering Perspective. *VaMoS*, 7, 17-26.
- LOUCOPOULOS, P.; STRATIGAKI, C.; DANESH, M. H.; BRAVOS, G.; ANAGNOSTOPOULOS, D.; DIMITRAKOPOULOS, G.; “Enterprise capability modeling: concepts, method, and application”; In *Enterprise Systems (ES), International Conference on* (pp. 66-77). IEEE, 2015.

- MAIA, L. G.; MORAES, M. M.; FREITAS, L. C. O.; “Elaboração e avaliação de modelo de gestão de pessoas orientado por competências”; *Perspectivas em Gestão & Conhecimento*, João Pessoa, v. 1, n. 1, p. 180-193, 2011.
- MCKEOWN, Max. *The strategy book*. Pearson UK, 2019.
- OMG; “Business Process Model and Notation (BPMN)”; Version 2.0, 2011.
- PENROSE, E.; “The Theory of Growth of the Firm”; London: Basil Blackwell, 1959.
- PICKTON, D. W.; WRIGHT, S. What's swot in strategic analysis?. *Strategic change*, 1998, 7.2: 101-109.
- PORTER M. E. *Competitive strategy*. New York: Free Press, 1980.
- PRAHALAD, C.K., HAMEL, G. The core competence of the corporation, *Harvard Business Review* (v. 68, no. 3) pp. 79–91, 1990.
- SAXENA, K.B.; “Capabilities versus Competence: How are they Different?”; disponível em : <https://www.linkedin.com/pulse/20141123155439-7430899-capabilities-versus-competence-how-are-they-different>.
- SOUSA, H.P.S.; LEITE, J.C.S.P. Modeling Organizational Alignment. *Lecture Notes in Computer Science*. 1ed.: Springer International Publishing, 2014, v. 8824, p. 407414.
- SOUSA, H.P., LEITE, J.C.S.P. Implementing GPI, a language for Organizational Alignment, Conceptual Modeling, Proceedings of the Eighth International i* Workshop (istar 2015), CEUR Vol-978, 2015.
- SOUSA, H.P.S.; LEITE, J.C.S.P. Toward an Organizational Alignment Modeling Language: The Human Resource Competency Perspective. In: 2017 IEEE 19th Conference on Business Informatics (CBI), 2017, Thessaloniki. 2017 IEEE 19th Conference on Business Informatics (CBI), 2017a. p. 277.
- SOUSA, H.P.S.; LEITE, J.C.S.P. Requirement Patterns for Organizational Modeling. In: 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), 2017, Lisbon. 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), 2017b. p. 252-259.
- SOUSA, H. P. S.; ALMENTERO, E. K.; LEITE, J. C. S. P.. Relacionando requisitos de software e competências de recursos humanos através de modelos organizacionais. In: Workshop on Requirements Engineering, 2019, Recife. Anais do WER19 - Workshop em Engenharia de Requisitos, 2019.
- STEVENS, M.J.; CAMPION, M.A.; "The knowledge, skill, and ability requirements for teamwork: Implications for human resource management"; *Journal of management* 20.2 (1994): 503-530.
- STIRNA, J.; GRABIS, J.; HENKEL, M.; ZDRAVKOVIC, J.; “Capability Driven Development – An Approach to Support Evolving Organizations”; In: Sandkuhl K., Seigerroth U., Stirna J. (eds) *The Practice of Enterprise Modeling*. PoEM 2012. Springer, Berlin, Heidelberg, 2012.
- STRATEGYZER, A. G. "Het Business Model Canvas." (2015).
- TRICHET, Francky; LECLÈRE, Michel. A framework for building competency-based systems dedicated to human resource management. In: *International Symposium on*

Methodologies for Intelligent Systems. Springer, Berlin, Heidelberg, 2003. p. 633-639.

VIGNOTTO, R.; “Como elaborar um (ótimo) dicionário de competências”; Portal Linked RH, 1ª edição, 2015.

Autores

Henrique Prado de Sá Sousa: é professor Adjunto do Departamento de Informática (DIA) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Doutor em Engenharia de Software pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Mestre em Engenharia de Software pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Bacharel em Sistemas de Informação pela Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Participou de diversos projetos de modelagem de processos de negócio e desenvolvimento de software, especialmente como engenheiro de requisitos. Ministra disciplinas da Engenharia de Software e Programação. Possui tese e dissertação no tema de construção de linguagens de modelagem para análise do alinhamento organizacional. Pesquisa continuamente a representação de conceitos específicos de domínio, orientado para a análise do alinhamento organizacional, bem como seu uso na engenharia de requisitos. Lattes: <http://lattes.cnpq.br/5825446300218282>.

Eduardo Kinder Almentero: É professor Adjunto do Departamento de Computação (DECOMP) da Universidade Federal Rural do Rio de Janeiro (UFRRJ). Doutor em Engenharia de Software pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Mestre em Engenharia de Software pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Bacharel em Ciência da Computação pela Universidade do Estado do Rio de Janeiro (UERJ). Tem experiência na área de Engenharia de Requisitos, atuando nas áreas de elicitação, modelagem e análise. Atualmente, ministra disciplinas de engenharia de software, gerência de projetos e gestão de TI. Realiza pesquisa sobre engenharia de requisitos, transparência de software e modelagem e alinhamento organizacional. Lattes: <http://lattes.cnpq.br/8851661551732110>.

Chapter

3

Transparência e Transformação Digital: O Uso da Técnica da Linguagem Simples

Claudia Cappelli, Vanessa Nunes e Rodrigo Oliveira

Abstract

This chapter presents a proposal for a short course on the use of Plain Language to support Transparency and the development of best practices for Digital Transformation. The objective of the short course is to introduce the concepts of Plain Language as a technique that can support the process of understanding the citizen and, consequently, Transparency and Digital Transformation, as a matter of public interest – right to citizenship. It aims to motivate participants to the challenges associated with (a) ensuring citizen understanding of public information, (b) providing Transparency in democratic contexts and (c) the use of emerging techniques that can help promote Transparency through Digital Transformation.

Resumo

Este capítulo apresenta o conteúdo utilizado para construção de um minicurso sobre a utilização da Linguagem Simples como suporte à Transparência e ao desenvolvimento das melhores práticas para Transformação Digital. O objetivo é apresentar uma introdução aos conceitos da Linguagem Simples (Plain Language) como uma técnica que pode apoiar o processo de entendimento do cidadão e consequentemente de Transparência e Transformação Digital, como uma questão de interesse público – direito à cidadania. O minicurso pretende motivar os participantes para os desafios associados: (a) a garantia do entendimento de informações; (b) ao fornecimento de Transparência em contextos democráticos; e (c) ao uso de técnicas emergentes que podem auxiliar à promoção da Transparência através da Transformação Digital.

3.1. Introdução

As áreas da vida humana moderna estão sendo absorvidas, aproximadas ou incorporadas ao digital [McNutt, 2014]. O uso de sistemas de informação para oferecer produtos e serviços é um processo sem volta. Sabemos que o mundo, cada vez mais e com mais velocidade, tem se tornado amplamente digital. Os benefícios vão desde rapidez no atendimento, passando por diminuição das distâncias até menores custos.¹

¹<https://www.gov.br/pt-br/noticias/financas-impostos-e-gestao-publica/2019/12/transformacao-digital-traz-economiza-de-tempo-e-recursos>

A digitalização no ambiente público se iniciou em 1957 quando o primeiro computador do Brasil foi comprado pelo governo do estado de São Paulo com o objetivo de modernizar o sistema de cobrança de contas do Departamento de Águas e Esgotos. O órgão, recém-criado, lidava com as dificuldades de abastecer a capital e outras cidades da região metropolitana, em um período em que a população urbana crescia em progressão geométrica.² Hoje, dos quase 4,5 mil serviços oferecidos pela administração pública brasileira para cidadãos e empresas, cerca de 3 mil já possuem trâmite totalmente digital e podem ser acessados pelo computador ou por *smartphones* (Brasil, 2021). No meio privado, não se sabe ao certo quando, mas essa mudança de paradigma se iniciou bem antes. O fato é que a digitalização em massa, não só no meio privado mas também no meio público, está impondo novos desafios à atuação das organizações. Os desafios se tornaram ainda mais evidentes com a aceleração deste processo imposta pela pandemia do vírus COVID-19.

A proliferação de tecnologias de acesso popular, engajamento social e mídias digitais revolucionaram a forma como as pessoas se comunicam, se informam, trabalham, fazem compras, realizam transações bancárias, participam da sociedade e interagem com os governos. Esse processo de digitalização demanda igualmente uma transformação digital nos serviços oferecidos por empresas, instituições e setores públicos. A Estratégia Brasileira de Transformação Digital (E-digital)³ visa proporcionar ferramentas para uma profunda transformação na atuação do próprio governo, na competitividade e produtividade das empresas, assim como na capacitação e inclusão na sociedade, para que todos possam se desenvolver e prosperar [Brasil, 2018]. Um dos eixos desta estratégia é em torno da Cidadania digital. A garantia do pleno exercício da cidadania no mundo digital. O objetivo é tornar o governo mais dinâmico, mais próximo da sociedade e mais eficiente facilitando a vida do cidadão. Para apoiar estas políticas públicas, existem três grandes linhas orientadoras, sendo o aumento da transparência; a ampliação da participação social em políticas públicas, e a criação de melhores serviços públicos em formato digital.

Não só no meio público mas também no privado, as organizações têm discutido, planejado e algumas já estão implantando e usando tecnologias para internet das coisas (IoT), conectividade, Big Data, Inteligência Artificial, computação em nuvem, gerando transformações nos seus processos de trabalho, cultura e relações.

O fato é que para garantir a introdução de grandes volumes de tecnologia da informação e comunicação, que sejam amplamente utilizadas pelas partes interessadas, de forma eficaz e eficiente, é necessário transformar essas ações em um pensamento estratégico que envolve reavaliar modelos de negócio, modelos produtivos e modelos de comunicação. É necessário transformar digitalmente uma organização e não somente digitalizar processos.

O termo Transformação Digital surge para propor discussões e abordagens que repensem os modelos organizacionais através da introdução de tecnologias da informação e comunicação (TIC). Este é um tema complexo, pela sua natureza

² <https://brazillab.org.br/noticias/exame-e-o-fim-da-burocracia-como-a-digitalizacao-de-servicos-publicos-avanca>

³ <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>

multidimensional (desde o nível estratégico até o chão de fábrica) e multidisciplinar (permeia todas as áreas de domínio das organizações). Por esta razão não existe uma definição comum e diferentes visões sobre o tema coexistem, nas quais uma ampla gama de conceitos, dimensões e dinâmicas concorrem e geram um panorama de pesquisa altamente fragmentado [Appio *et al.*, 2021]. Uma coisa que eles têm em comum: é necessário transformar as TICs em vantagem competitiva (ou benefício público). Os desafios são inúmeros, em diversas áreas de domínio, como era de se esperar.

O desafio que queremos abordar nessa discussão está no fato de que escalar o acesso a dados e informações, a solicitação e acompanhamento de serviços e as relações de comunicação de forma digital e (na maioria das vezes) assíncrona tem requerido crescente transparência das informações e do seu fluxo de transformação. O fato é que a transparência, ou a falta dela, tem estado no topo das agendas públicas e também das privadas. Sua importância tem sido realçada com a crescente demanda por serviços digitais assíncronos, e o holofote tem sido colocado principalmente na oferta de serviços públicos através de software. Este é, inclusive, um dos objetivos do E-Digital. Claro que transparência organizacional (ou no ambiente público, a chamada transparência pública) é uma demanda anterior ao pensamento de transformação digital das organizações, como, por exemplo, a publicação da Lei 12.527, chamada Lei de Acesso à Informação, datada de 18 de novembro de 2011. O que acontece é que todo esse processo tecnológico que permitiu aumentar e disseminar mais rapidamente o acesso a dados e informações colocou um holofote sobre o tema e pontuou a necessidade de acelerar a implantação de mecanismos que o garanta.

Transparência pode ser definida como a característica que possibilita ao usuário acessar, entender e ter facilidade e qualidade de uso destas informações [Cappelli; Leite, 2008]. O acesso à informação permite criar uma sociedade democrática com cidadãos participativos, fornecendo-lhes ferramentas para compreender e utilizar a informação e estimulando um pensamento crítico sobre as informações e serviços prestados. Um ponto positivo é que temos visto disponível no contexto público uma abundância de portais de serviços, dados e informações geradas continuamente sobre finanças, saúde, desenvolvimento humano, entre outros [Barcellos, 2017].

Contudo, tais informações públicas são de pouco ou nenhum proveito a menos que os cidadãos entendam o que elas significam [Song; Lee, 2016]. Esse entendimento se faz ainda mais necessário, visto que o governo é o agente que regula a vida social [Layne; Lee, 2001]. E, na prática, apenas um pequeno número de cidadãos se torna capaz de entender dados e informações e fazer uso deles [Park; Gil-Garcia, 2017]. Então, apesar de possibilitar o acesso ser um dos focos principais da agenda pública, este é apenas o primeiro passo, mas, não o único, para que uma informação alcance a transparência.

De fato, como aponta a definição de transparência, para as pessoas usarem efetivamente serviços e se apropriarem de dados e informações que ajudem a fomentar uma participação ativa, principalmente no ambiente público, deve além de dar acesso às informações precisam ser apresentadas de forma que possam ser entendidas. Acontece que no Brasil, um dos desafios é o alto nível de brasileiros iletrados e desfavorecidos

digitalmente com difícil compreensão das informações. Segundo o Indicador de Analfabetismo Funcional [INAF, 2018], 63% da população brasileira está abaixo da classificação proficiente ou intermediária para a competência de entendimento em informações simples. A cada 10 brasileiros, 3 têm muita dificuldade para usar a leitura e a escrita, comparar ou avaliar informações.

Uma técnica que favorece o entendimento para uma efetiva transparência, e com isso é uma ferramenta de apoio à transformação digital: a Linguagem Simples (Plain Language)⁴. Essa linguagem faz parte de um movimento cujo foco principal é fazer com que o leitor encontre rapidamente a informação que precisa, consiga entender o que foi encontrado e faça uso dela com facilidade. Ela surgiu em sociedades de língua inglesa na década de 70. Ganhou impulso porque teve apoio de associações de defesa do consumidor, funcionários públicos e pessoas da área do Direito insatisfeitas com uma comunicação complexa. Mesmo não tendo um padrão no início, o movimento foi se espalhando devido à percepção de sua importância.

Percebemos nesse movimento a discussão sobre como oferecer mecanismos para ajudar as pessoas a entender melhor as informações e com isso cumprir de modo mais fácil e rápido suas tarefas ou utilizar serviços (públicos e privados) corretamente [Barboza, 2010]. A Linguagem Simples apoia os leitores a entender o mundo ao seu redor, especialmente aqueles com pouca (ou nenhuma) alfabetização. E isso se dá através da adaptação de documentos, gráficos, figuras, modelos de processos, modelos organizacionais, websites e interfaces de sistemas.

Nosso propósito neste capítulo é: introduzir os conceitos de Transformação Digital, Transparência, Entendimento e Linguagem Simples mostrando o forte relacionamento entre eles; Fornecer um panorama das principais ações em Linguagem Simples; e Apresentar as diretrizes da Linguagem Simples, definidas do *Federal Plain Language Guidelines*⁵.

Este capítulo está organizado da seguinte forma: A Seção 3.2 apresenta os fundamentos teóricos do tema. Em seguida destacamos as principais ações na Seção 3.3. Uma descrição das práticas apresentadas no *Federal Plain Language Guidelines* consta na Seção 3.4. Concluimos na Seção 3.5, seguido das referências utilizadas neste capítulo.

3.2. Fundamentação Teórica

3.2.1. Transformação Digital

Transformações Organizacionais são constantes e necessárias para evolução dos negócios privados e públicos. E a capacidade de se transformar e se adaptar às novas demandas determina quais organizações vão se destacar. Este tema não é novidade quando discutido pela comunidade que trabalha com os temas de gestão e administração, porém a introdução, quase que obrigatória, do elemento TIC neste contexto, colocou uma nova variável neste ambiente conhecidamente desafiador que

⁴ <https://plainlanguagenetwork.org/plain-language/what-is-plain-language/>

⁵ <https://www.plainlanguage.gov/guidelines/>

precisa ser equacionada. O avanço crescente das novas tecnologias digitais, o dinamismo dos mercados e a demanda por se tornar presente em todos os lugares e para todos, torna crítica a consideração de TIC como elemento transformador de modelos de negócio e não “mero” mecanismo de digitalização de dados, informações e formulários.

Neste sentido, as ações de TIC deixam de ser vistas como processos de suporte às organizações para serem consideradas nas discussões estratégicas [Collins *et al.*, 2014] de avaliação de mercado, direcionamento, definição de táticas operacionais e como as TIC podem potencializar e inclusive definir como os modelos negociais, de comunicação e vendas serão desenvolvidos e implantados.

Claro que este tema tão complexo, multidimensional e interdisciplinar gera uma pluralidade de definições, abordagens e modelos. De fato, houve um aumento significativo de publicações acadêmicas e de mercado já percebida por Henriette *et al.* (2015). Diversos autores acadêmicos discutem e evoluem definições sobre o tema, desde uma visão puramente técnica/tecnológica [Resnick, 2002] até as definições que incorporam pensamento estratégico e de transformação do negócio através de aumento da capacidade digital [Sandegberg *et al.*, 2014] [Kane, 2017] [Morakanyane *et al.*, 2017] [Macalintal; Chepkasova, 2017].

Os quatro fatores-chave de sucesso que Laudon e Laudon (2016) apontam como base para conduzir a Transformação Digital nas organizações são as pessoas, os processos, as tecnologias e a cultura. Indo mais além, de acordo com Rogers (2016) existem cinco domínios associados à transformação digital que devem ser observados e transformados de forma integrada: os (tipos de) clientes que a organização tem e quer ter; os níveis de concorrência e cooperação que uma organização tem e quer ter com outras de seu mesmo nicho; a forma como a organização trata e manipula dados e informações; a forma como a organização promove inovação; e a proposta de valor que a organização oferece aos seus clientes.

Explorando a combinação desses elementos estudados e pensando no direcionamento de ações para promover a Transformação Digital nas organizações, segundo estudo realizado por Tadeu *et al.* (2018), os autores identificaram dez dimensões que podem ser consideradas dentro do contexto da transformação digital:

- **Estratégia Digital:** as organizações precisam definir os objetivos digitais que foquem nas transformações dos seus negócios e pensar neles de forma integral e não em ações isoladas.
- **Tecnologias Digitais:** as organizações precisam analisar e decidir sobre quais tecnologias utilizar para implementar suas estratégias digitais e gerar valor. Não escolha IoT, IA, Big Data, 3D, etc sem antes avaliar que valor podem gerar para os seus negócios.
- **Capacidade Analítica:** as organizações precisam desenvolver forte capacidade analítica, e em saber tratar de forma inteligente seus dados e informações para obter ganhos e gerar valor. Aqui falamos de organização, entendimento, inteligência, disponibilidade e transparência.
- **Relacionamento com clientes:** as organizações precisam migrar suas formas de comunicação, desde o primeiro contato ou pré-venda até o relacionamento pós-venda. E as novas TIC têm o poder de ajudar as empresas a transformar e

melhorar essa comunicação e interagir de forma mais inteligente e interessante com o cliente/cidadão.

- Relacionamento em rede: as organizações precisam estender suas ações de transformação digital aos seus parceiros de negócio, sejam fornecedores, governos, universidades, startups, fundos de investimento, etc. Essa integração e transparência desenvolvem cadeias mais responsivas e eficientes.
- Estrutura organizacional e processos digitais: as organizações precisam reformular toda sua estrutura organizacional e processos de trabalho, em todos os níveis, para entrar de cabeça na transformação digital, a começar por aplicar internamente todas as transformações digitais que envolvem automação de processos, comunicações, análise de dados, transparência interna e estruturas flexíveis voltadas para geração de valor através do uso de recursos digitais.
- Cultura e pessoas: a transformação digital demanda grandes mudanças culturais no sentido de promover e demandar mais colaboração, orientação a inovação e decisões orientadas por dados e fatos. Isso requer revisão da formação das pessoas para essa mudança de paradigma, provimento de fluência digital e forte liderança para criar e fortalecer essa nova cultura digital.
- Riscos e investimentos: a transformação digital oferece mecanismos robustos de suporte à gestão de riscos baseados em profunda e sólida análise contínua de dados. Isso abre espaço para promoção de inovações mais ousadas, porém pensadas em cima de decisões melhor fundamentadas e aumenta a segurança dos investidores.
- Aspectos legais e éticos: as organizações precisam entender quais mudanças de paradigma a transformação digital e as novas TIC trazem para as mudanças dos modelos organizacionais e os possíveis impactos legais que isso pode trazer e se precaver.
- Novos modelos de negócios digitais: uma vez iniciada a bem conduzida a transformação digital novos modelos de negócio digitais podem surgir. As organizações, através de mecanismos de inovação, devem estar atentas e buscando oportunidade que as novas TIC podem trazer para expandir e (re)transformar seus negócios.

Em Dias (2019), o autor faz uma breve discussão de seis *frameworks* que propõem caminhos a serem conduzidos para uma ou mais das dimensões citadas.

3.2.2. A Lei do Governo Digital

Em 29 de março de 2021 foi publicada a Lei 14.129 que “dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública”. Esta lei estabelece regras e instrumentos para o aumento da eficiência da Administração Pública, especialmente por meio da inovação, da transformação digital e da participação dos cidadãos.

Segundo o Ministério da Economia⁶, “Com o objetivo de ampliar a oferta de soluções digitais e facilitar a vida dos brasileiros, a lei institui, entre outras diretrizes:

⁶<https://www.gov.br/economia/pt-br/assuntos/noticias/2021/junho/lei-do-governo-digital-entra-em-vigor-com-foco-na-eficiencia-e-inovacao>

serviços digitais acessíveis por dispositivos móveis (como o aplicativo Meu INSS e a Carteira de Trabalho Digital); uso de plataforma única de acesso a informações e serviços, o gov.br; estímulo às assinaturas eletrônicas nas interações entre órgãos públicos e cidadãos (assinatura avançada nas juntas comerciais, por exemplo); fortalecimento da transparência e do uso de dados abertos pelo governo; além da aplicação da tecnologia para otimizar processos de trabalho da Administração Pública”.

Um dos grandes destaques da lei é o estabelecimento de uma plataforma única de acesso pelo cidadão, que por si só, já traz diversos dos desafios de transformação digital já citados e estão sendo enfrentados pelos diversos órgãos de governos em todos os poderes e esferas.

Ao falar de transparência e linguagem simples destacamos alguns trechos que a lei aborda e demanda no Artigo 2º:

- “simplificação da relação do poder público com a sociedade”
- “a disponibilização em plataforma única do acesso às informações e aos serviços públicos”
- “a transparência na execução dos serviços públicos e o monitoramento da qualidade desses serviços”
- “o incentivo à participação social no controle e na fiscalização da administração pública”
- “o uso de linguagem clara e compreensível a qualquer cidadão”
- “a simplificação dos procedimentos de solicitação, oferta e acompanhamento dos serviços públicos, com foco na universalização do acesso e no autosserviço”
- “a promoção de dados abertos”

Estes trechos deixam claro, mais uma vez, como a transformação digital requer a implantação de mecanismos de transparência e comunicação clara e simples com os cidadãos brasileiros. Em suma, a Lei Do Governo Digital representa um importante passo para a universalização do acesso aos serviços públicos e para criar o conceito de governo como um todo e para todos.

3.2.3. Transparência

A ideia de "transparência" possui múltiplos sentidos em diversas áreas. Cappelli (2009) agregaram várias características que compõem o termo para o contexto organizacional. Aspectos como a acessibilidade, usabilidade, entendimento e auditabilidade agregam clareza às informações. Com isso, definiram-se os denominados degraus para a transparência, conforme a Figura 1, um modo de institucionalizar esse conjunto de características associadas em produtos e serviços.

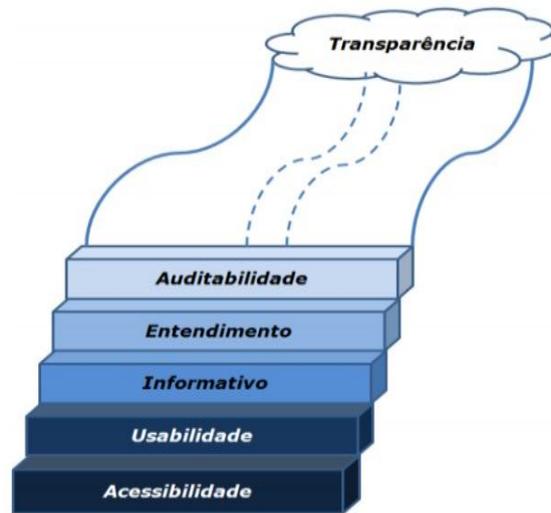


Figura 3.1. Degraus da Transparência [Cappelli, 2009]

A construção do conceito de transparência é feito através de requisitos que se relacionam, atuando como degraus para um nível mais elevado. Em outros termos, para termos uma informação transparente é preciso que anteriormente ela seja entendível. De igual forma para ser entendida deve estar acessível e da mesma maneira que sua representação seja simples, fácil e assim sucessivamente. Todas as interações entre os aspectos favorecem ou influenciam na obtenção da transparência total. Mesmo que, na prática, não haja impedimento de determinada organização atingir pontos de um nível superior dos degraus deixando de cumprir outros de um grau anterior.

Essa estrutura de influência de características de qualidade é mapeada pelos autores em um grafo, conforme a Figura 2. Nele se estende cada conceito em ações a serem implantadas nas organizações conforme o nível de satisfação desejado. Como resultado, temos a formalização do conceito de transparência da informação de acordo com Cappelli (2009) como a característica que possibilita ao cidadão não só acessar, mas obter facilidade de uso, qualidade de conteúdo, entendimento e auditoria às informações de seu interesse, sob a tutela de centros de autoridade.

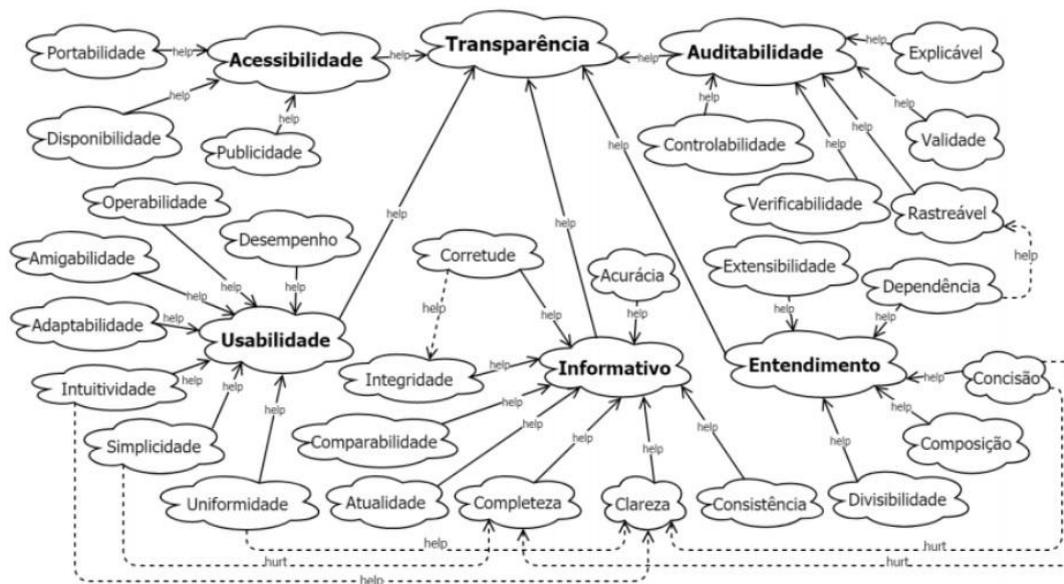


Figura 3.2. Grafo de características da Transparência. [Cappelli, 2009]

3.2.4. Entendimento

Esta característica na definição de transparência de Cappelli (2009) se apresenta como um dos pontos mais importantes no trato com o cidadão. Apesar de diversas leis se preocuparem bastante com dar acesso, e garantir o uso, além claro de que os dados tenham qualidade, não se percebe ainda um estímulo a garantir que as informações apresentadas sejam de fato entendidas pela sociedade. O que nos leva a outro problema dado que se o cidadão não entende ele não faz uso e conseqüentemente não participa de discussões e decisões. O nível do entendimento é uma característica, que segundo Cappelli e Leite (2008) se refere à capacidade de alcançar o significado e o sentido. Ainda que os demais aspectos sejam necessários para o cidadão ter transparência de informações, o entendimento das informações é que fará com que ele de fato se aproprie das mesmas e consiga fazer uso. Para isso precisamos cada vez mais de mecanismos que propiciem este entendimento e a Linguagem Simples é um destes.

3.2.5. Linguagem Simples

Linguagem simples é uma das traduções do termo em inglês “Plain Language”. Podemos ainda traduzir como Linguagem Clara ou Linguagem Cidadã. O conceito surgiu devido à necessidade de facilitar o entendimento da informação, principalmente aquelas que as organizações apresentam à sociedade. Na década de 1970, um movimento mundial foi iniciado de modo a consolidar orientações para a escrita e a organização visual da informação. Desde então, a prática de Linguagem Simples se disseminou por diversos países. Ao longo deste tempo também nasceram algumas associações e uma Federação que criaram juntas o *Federal Plain Language Guidelines*,

que contém um corpo de conhecimento sobre as melhores práticas indicadas pela Linguagem Simples. Neste são apresentadas diretrizes e recomendações para apoiar uma escrita com mais clareza e objetividade. Esta Federação definiu também três princípios para esta Linguagem, indicando que uma comunicação está em linguagem simples quando a escrita, a organização, o design e todo e qualquer outro elemento utilizado para apresentar informação sejam tão claros que o público-alvo consiga encontrar facilmente o que procura, compreender o que encontrou e usar essa informação. Seguimos com essa declaração pela sua abrangência, conceituando o objetivo da linguagem na comunicação de forma ampla, não apenas textual. Apesar da escrita ser o principal meio de aplicação da técnica.

A lei da redação simples (*Plain Writing Act*) [Gov Info, 2010] nos Estados Unidos sancionada em 2010 estabelecendo que os documentos governamentais emitidos ao público fossem redigidos de forma clara e a política de comunicação do governo canadense que institui também o uso da Linguagem Simples são iniciativas proeminentes na parte norte-americana. As ações oficiais da União Europeia são encontradas no serviço de publicações como a distribuição de manuais de como escrever claramente. Na América Latina o governo colombiano desponta com um Programa Nacional de Serviços ao Cidadão desde 2013 estabelecendo a Linguagem Clara como uma das prioridades da administração pública. Guias, métodos e seminários são divulgados e já foram traduzidos mais de 150 documentos de alto impacto.

Já as iniciativas brasileiras têm referências explícitas sobre a aplicação da Linguagem Simples em websites governamentais brasileiros. A rigor, não existe uma diretriz definida sobre este tema a não ser por citações à Lei de Acesso à Informação, nos respectivos e obrigatórios portais de transparência. O estado de São Paulo vai à vanguarda divulgando ações para elaboração de uma linguagem mais clara para tradução de suas informações serem mais acessíveis à compreensão do cidadão comum. Contudo, independente dos esforços, Fischer *et al.* (2019) demonstram que textos essenciais de benefícios do governo federal tem alto grau de dificuldade para leitura seja pela análise textual manual ou automatizada com métricas geradas por ferramentas como o Coh-Matrix-Port 2.0 que utiliza de lógica para criar/identificar o coeficiente de dificuldade do texto como, por exemplo:

- Índice de legibilidade – correlação entre tamanhos médios de palavras, sentenças e a facilidade de leitura
- Sílabas por palavra de conteúdo – número médio de sílabas em substantivos, verbos, adjetivos e advérbios
- Modificadores por sintagma – média de número de advérbios, adjetivos e artigos que participam de um sintagma.

A Linguagem Simples é projetada para comunicar precisamente as informações para o público pretendido, com documentos claros e concisos [Barboza, 2010]. A organização norte-americana FDA (*Food and Drug Administration*) apresenta muitos benefícios que a experiência com a Linguagem Simples oferece, entre eles: (i) Melhorar a compreensão pública das comunicações do governo; (ii) Economizar dinheiro e aumentar a eficiência; (iii) Reduzir a necessidade de esclarecimento público; (iv)

Diminuir os recursos gastos com fiscalização; (v) Minimizar erros cometidos e o esforço para corrigi-los.

Shumskaya (2019) compila uma série de argumentos contra ideias de que a Linguagem Simples é apenas um conjunto de regras engessadas para tornar um texto mais curto e infantil e de que não existem provas concretas de sua eficácia. O trabalho destaca com veemência a linguagem como uma ferramenta importante de capacitação dos cidadãos para tomar decisões conscientes e informadas. Reforçando sua eficácia e garantia de melhorias na comunicação com o seu uso.

Apesar de toda a sua contribuição, ainda existem muitas oportunidades de trabalho a partir deste conhecimento disponibilizado. Quanto à dinâmica escrita, a Linguagem Simples oferece métodos para melhorar a inteligibilidade, entretanto no carácter visual ainda pode ser bastante aprofundada. A partir do que diz a própria definição da linguagem, que além da escrita do texto, a estrutura e o design também devem se apresentar de forma melhor compreensível, pouco se encontra no corpo da própria linguagem sobre o uso de elementos visuais. Estes outros recursos se tornam muito importantes por serem amplamente integrados nas comunicações, justamente pela sua fácil compreensão e carácter atrativo. Identificamos também que a mesma pode se aprofundar a oferecer método que abrange o tópico de visualização, como uma das áreas de fundamentação teórica para o uso de imagens para o entendimento de dados e informações [Oliveira, *et al.*, 2021].

Outro ponto importante que trazemos são as formas como esta linguagem pode ser aplicada. Citamos aqui 4 usos para a mesma:

- Adaptação de textos, modelos de processo, modelos de regras, planejamentos e objetivos, documentos, websites, interfaces de sistemas.
- Organização de informação em repositórios, metadados, modelos conceituais.
- Treinamentos e Mentoria para transferência de conhecimento.
- Adaptação e Desenvolvimento de ferramentas para suporte e automação do uso.

3.3. Principais Ações em Linguagem Simples

Na Linguagem Simples a relevância dos resultados é demonstrada pelas inúmeras iniciativas, públicas e privadas, com intuito de fomentar e aplicar o uso desta linguagem no Brasil e no Mundo. Destacam-se:

3.3.1. Ações Internacionais

- Plain Language Association International⁷ (PLAIN) é uma organização canadense sem fins lucrativos que reúne defensores da linguagem simples e profissionais de todo o mundo em mais de 30 países que trabalham em comunicação clara em pelo menos 15 idiomas.
- Clarity International⁸ é uma organização britânica sem fins lucrativos que reúne profissionais empenhados em promover uma linguagem jurídica simples. Possui representantes oficiais em cerca de 50 países que defendem o uso de linguagem jurídica simples no lugar do “juridiquês”.
- Center for Plain Language⁹ é uma organização norte-americana composta por profissionais de diversas áreas como acadêmicos, consultores, organizações de saúde e a comunidade de negócios com a missão de criar uma cultura de clareza onde com uma comunicação clara as pessoas e organizações possam prosperar.
- International Plain Language Federation¹⁰ coordena as ações conjuntas das três organizações citadas anteriormente para promover o uso e a prática da linguagem simples.
- The Plain Language Action and Information Network¹¹ é uma plataforma mantida por funcionários do governo norte-americano. Seu objetivo é difundir o uso da linguagem simples em agências e órgãos estatais.
- Plain English Campaign¹² é uma organização britânica que oferece serviços de edição e análise de textos para linguagem simples e é responsável pelo selo Crystal Mark para documentos de compreensão fácil.
- Plain Language Commission¹³ é uma organização britânica que oferece serviços de edição de textos em linguagem simples e cursos no tema. Criou a certificação Clear English Standard.
- Plain English Foundation¹⁴ é uma fundação australiana que oferece consultorias e treinamentos em linguagem simples.
- Simplification Centre¹⁵ é uma organização britânica sem fins lucrativos que promove linguagem simples com foco principal nas áreas da saúde, direito e governo.

⁷ <https://plainlanguagenetwork.org/>

⁸ <http://www.clarity-international.org/>

⁹ <https://centerforplainlanguage.org/>

¹⁰ <https://www.iplfederation.org/>

¹¹ <https://www.plainlanguage.gov/>

¹² <http://www.plainenglish.co.uk/>

¹³ <https://www.clearest.co.uk/>

¹⁴ <https://www.clearest.co.uk/about>

¹⁵ <https://www.simplificationcentre.org.uk/>

- Red de Lenguaje Claro¹⁶ é uma organização chilena que reúne 7 instituições públicas com o objetivo de trabalhar em conjunto na implementação de ações que visem gerar iniciativas, projetos e medidas que promovam, divulguem e facilitem o uso da linguagem clara.

3.3.2. Ações Nacionais

- Guia de Edição de Serviços do Gov.br¹⁷ é um guia desenvolvido pelo Governo Federal sobre como melhorar o texto de um serviço utilizando linguagem simples. Ele contém diretrizes e recomendações para o uso de linguagem simples para editores do portal do governo federal, auxiliando a manter os textos mais claros e úteis para os usuários desses serviços.
- Projeto Acessibilidade TT¹⁸ é um Projeto de pesquisa da Universidade Federal do Rio Grande do Sul que aborda a acessibilidade textual e terminológica
- O Laboratório Interdisciplinar de Linguagem Cidadã¹⁹ sediado no Rio de Janeiro, congrega pesquisadores atuando em pesquisa teórica e aplicada sobre linguagem cidadã.
- O Programa Municipal de Linguagem Simples da prefeitura de São Paulo²⁰, criado em 2019, quer simplificar a linguagem que a Prefeitura de São Paulo usa na comunicação com a população.
- As orientações para Adoção de Linguagem Clara no Portal Governo Aberto SP²¹, criadas em 2016, oferecem um guia que é parte integrante do projeto de cooperação entre o Governo do Estado de São Paulo e o Reino Unido.
- O Laboratório de Inovação e Dados do Governo do Estado do Ceará²² (Iris Lab Gov) foi criado com a proposta de fomentar e ampliar a inovação no setor público que tem como um de objetos de pesquisa e investimento a discussão e implantação da linguagem simples no Estado.

¹⁶ <http://www.lenguajeclarochile.cl/>

¹⁷ <https://www.gov.br/pt-br/guia-de-edicao-de-servicos-do-gov.br/escrevendo-para-o-seu-usuario>

¹⁸ <http://www.ufrgs.br/textecc/acessibilidade/>

¹⁹ <https://linclab.com.br/>

²⁰ <https://011lab.prefeitura.sp.gov.br/linguagem-simples/inicio>

²¹

http://www.governoaberto.sp.gov.br/wp-content/uploads/2017/12/orientacoes_para_adocao_linguagem_clara_ptBR.pdf

²² <https://www.egp.ce.gov.br/irislabgov/>

3.3.2. Leis Nacionais

- Lei de Acesso à Informação (Lei 12.527 de 18 de novembro de 2011), determina que todo órgão e entidade pública ofereça o acesso às suas informações utilizando de procedimentos objetivos, ágeis, transparentes, claros e em linguagem de fácil compreensão.
- Lei do Governo Digital (Lei 14.129 de 29 de março de 2021) dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública.
- Lei 13.460 de 26 de junho de 2017 dispõe sobre participação, proteção e defesa dos direitos do usuário dos serviços públicos da administração pública. Esta lei regulamenta a Carta de Serviços, onde cada município deverá elaborar a sua carta com informações claras e precisas, tendo que dispor em locais de fácil acesso, tais como o portal institucional e o local onde os serviços serão prestados.

3.4. Práticas para implementar a Linguagem Simples

Dado que o objetivo principal da Linguagem Simples é facilitar a compreensão, a localização e o uso das informações, podemos dizer que certamente ela é base para a transparência e conseqüentemente o ponto inicial para a participação cidadã. No caso da transparência podemos dizer ainda que garante que a sociedade acesse facilmente os dados emitidos pelos órgãos do governo e acompanhe, de modo mais atuante, a gestão pública. Somente a clareza e conhecimento das informações permite um cidadão engajado e participativo.

Para isso faz-se necessário que as organizações, principalmente as da esfera pública, estejam preparadas para o uso desta linguagem. Textos com jargões e termos técnicos, elaborados de maneira complexa e com palavras de difícil entendimento apenas geram desinteresse e confusão. Muitas vezes o cidadão precisa de uma terceira pessoa que possa lhe explicar o que está sendo dito com aquela determinada informação. Junte-se a isso a tradicional forma de escrita dos governos que são extremamente burocráticas, enfraquecendo a comunicação entre o poder público e os cidadãos que ficam muitas vezes sem entender seus direitos e deveres apesar de isso ser um direito constitucional.

Deve-se considerar que uma Linguagem Simples também torna a comunicação mais inclusiva, o que promove acessibilidade. Para entender melhor as práticas da Linguagem Simples e poder aplicá-las é importante ter domínio do conteúdo apresentado no *Federal Plain Language Guidelines*. Este guia, como já citado anteriormente, está sob a tutela de uma federação sendo organizado e mantido por organizações ligadas a esta federação. Este guia é estruturado em 5 grandes blocos de conteúdo como apresentado na Figura 3. Neste trabalho detalharemos e mostraremos através de alguns exemplos as principais práticas definidas nos três primeiros blocos.

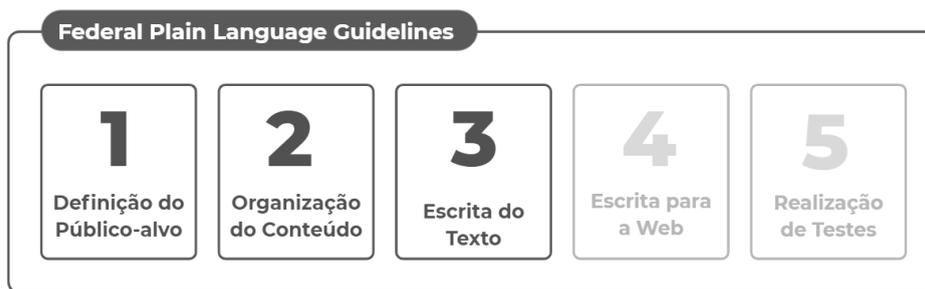


Figura 3.3. Conteúdo do *Federal Plain Language Guidelines*. Fonte: Os Autores.

3.4.1. Definição do Público-alvo

Esta prática indica que devemos buscar conhecer o público alvo da informação que estamos tentando transmitir, entender o quanto ele já sabe sobre o tema, o que mais ele precisa saber e que perguntas ele faria sobre aquele conjunto de informações. Além disso, é importante pensar qual será o melhor resultado a ser apresentado e como devo me expressar para apresentar este resultado dessa forma. Esta prática se divide em duas diretrizes: como escrever para sua audiência e como tratar diferentes audiências.

3.4.1.1. Como Escrever para sua Audiência

A técnica de definir Personas pode ajudar neste caso. Persona é a representação fictícia do seu cliente típico. Ela é baseada em: dados reais sobre o comportamento, características demográficas, histórias pessoais, motivações, objetivos, desafios e preocupações. A Figura 4 apresenta um exemplo e a Tabela 1 apresenta um exemplo de uso da Linguagem Simples adaptada a uma possível persona definida.

Exemplo de Persona

Paulo Correa – Técnico de Suporte

Paulo Correa, de 43 anos, trabalhou durante muitos anos consertando e configurando computadores. Atualmente, trabalha na universidade AprendaMais, configurando PCs e as contas de alunos de cada turma. Ele faz um curso de administração de redes, mas prefere aprender fazendo do que assistindo aulas ou lendo manuais. Quando tem alguma dúvida, ele faz uma busca na Internet por informações que lhe ajudem a resolver os seus problemas. Usuário "das antigas", Paulo prefere utilizar linguagem de comando do que assistentes em interface gráfica, pois acredita que assim seja mais eficiente. Sempre que uma tarefa se repete com frequência, ele tenta elaborar um script ou fazer uma configuração que acelere seu trabalho.

Todo início de período, Paulo precisa configurar dezenas de contas para cada turma, com diferentes perfis, fornecendo acesso diferenciado para alunos regulares, monitores, instrutores e coordenadores de cada disciplina. Precisa atender aos pedidos dos professores sobre o que deve estar disponível na intranet de cada disciplina (e.g., material didático, fórum de discussão, recebimento de trabalhos de alunos, cadastramento de notas, pedidos de revisão). Seu maior objetivo é atender aos professores com a maior eficiência possível. Para isso, é importante ele poder acessar o sistema onde quer que esteja, no horário que for, para realizar qualquer tarefa remotamente.

Adaptado de Barbosa e Silva (2010)

Figura 3.4. Exemplo de Persona Adaptado de Barbosa e Silva (2010)

Tabela 3.1. Exemplo de uso de Persona

Texto Comum	Texto com Linguagem Simples
“O atendimento para solicitação de desconto no IPTU residencial será disponibilizado nas agências regionais ao público em dias úteis em horário comercial”	“Se você é dono de uma casa na cidade de Patópolis e ganha até 5 salários mínimos, pode solicitar desconto do IPTU em qualquer agência municipal de segunda a sexta das 09h às 18h”

3.4.1.2. Tratar diferentes audiências

A depender da quantidade de personas definidas como público-alvo da sua informação, pode ser importante abordar públicos diferentes, separadamente. Muitos conteúdos abordam mais de um público. Conjuntos de informação que misturam elementos destinados a públicos diferentes podem confundir os leitores, pois, ao abordar diferentes públicos no mesmo lugar, você dificulta para cada público encontrar o material que se aplica a eles dificultando que cada público possa cumprir os requisitos para sua necessidade. neste caso o mais indicado é que se usem formas de apresentação de informação ou até mesmo conteúdos diferenciados para as personas ou grupos de personas definidos.

3.4.2. Organização do Conteúdo

Organizar conteúdos na maioria das vezes é a chave para o entendimento. Em geral, lemos documentos, analisamos informações ou visitamos sites para obter respostas para nossos problemas ou informações para atender nossas necessidades. Queremos saber como fazer algo ou o que acontece se não fizermos nada. Também queremos obter esse conhecimento o mais rápido possível. Para tal o *Federal Plain Language Guidelines* sugere algumas práticas para fazer melhor esta organização. Ele aborda 4 diretrizes: atender às necessidades da audiência, falar para uma pessoa e não para um grupo, usar títulos úteis e escrever seções curtas.

3.4.2.1. Atender às necessidades da audiência

Para todo conteúdo que se queira transmitir é importante começar sempre declarando a finalidade do documento e seus resultados. Deve-se eliminar qualquer preenchimento e conteúdo desnecessário e atentar para colocar as informações mais importantes no início da apresentação. Informações básicas (quando necessário) devem ser colocadas no final. Outro ponto bastante importante é que o seu conteúdo deve estar organizado para responder às perguntas que o seu público provavelmente fará. Busque estruturar na ordem em que ele as faria.

Para conjuntos de informação, mas complexos, podem ser criados índices que facilitem a navegação ou a busca direta. No caso de sites, os menus podem também ser um grande aliado para isso. Estes mecanismos ajudam os usuários a seguir e encontrar

rapidamente as informações de que precisam. Porém, não crie muitos níveis de hierarquia. No caso de documentos recomenda-se que seja no máximo 3.

No caso de um processo ou um serviço, pode-se apresentar as etapas dos mesmos na ordem em vão ser usadas ou executadas. Isso não só facilita o entendimento como também guia o usuário diminuindo a incidência de erros.

Já considerando um conteúdo que possui informações gerais e específicas, coloque as informações gerais em primeiro lugar e as especializadas ou exceções depois. O que se dirige à maioria dos leitores na maioria das situações deve vir em primeiro lugar.

Se existir mais de um público para a sua informação, procure abordar cada um separadamente. Não é agradável ter que olhar informações destinadas a outro público. Personalização é uma técnica que pode e deve ser utilizada sempre.

3.4.2.2. Fale para uma pessoa, não para um grupo

Um fator bastante importante é falar com quem está lendo a informação. Mesmo que sua comunicação vá atingir um milhão de pessoas, lembre que você está falando com a única pessoa que o está lendo. Não dar atenção a isso pode gerar confusão se uma determinada ação que está sendo descrita deve ser realizada pela pessoa que lê ou por um grupo. Uma boa prática bastante recomendada é usar “você” para se dirigir ao usuário. O uso do “ele ou ela” pode fazer com que o leitor não perceba que ele é o agente da ação. A tabela 2 ilustra como fazer uso desta prática.

Tabela 3.2. Exemplo de uso de fala direta

Texto Comum	Texto com Linguagem Simples
Os residentes de Patópolis que buscam assistência no processo de preenchimento e envio de sua inscrição podem entrar em contato com a agência usando o número de telefone listado abaixo.	Se você precisar de ajuda com sua inscrição, ligue para a agência usando o número abaixo.

3.4.2.3. Use Títulos úteis

Uma maneira eficaz de facilitar o entendimento de um conjunto de informações, um sistema, um documento ou até um site é usar títulos auto explicativos e úteis. Existem algumas categorias de títulos:

- **Título de perguntas:** Muito úteis, mas é necessário perceber quais perguntas o seu público fará. Se essas perguntas são possivelmente conhecidas, use-as como títulos. Elas ajudarão o público a encontrar rapidamente as informações que procuram. Um bom exemplo disso são os FAQs (*Frequently Asked Questions*)

que atualmente a maioria das organizações disponibiliza em seus sites reunindo as perguntas mais frequentes feitas e as respostas das mesmas.

- **Título de declarações:** Define em poucas palavras o que se segue no conteúdo. É sempre uma ótima escolha porque resume bem o que segue.
- **Título de Tópicos:** São do tipo “Geral”, “Aplicativo” e “Escopo”. Devemos evitar sempre que possível, pois são vagos e não ajudam a saber o que tem de fato no que segue.

Estruturar primeiramente os títulos seja de um texto, um conjunto de dados ou de um menu pode nos ajudar a perceber a estrutura, verificar se não faltou nada e também nos mostrar possíveis agrupamentos. Em todos os casos citados, os títulos nunca devem ser muito longos. Um máximo de 5 palavras pode ser uma boa prática. A Tabela 3 mostra um exemplo.

Tabela 3.3. Exemplo de títulos úteis

Texto Comum	Texto com Linguagem Simples
Levantamento de Profissionais da saúde não vacinados contra COVID-19	Prefeitura realiza esta semana um levantamento de trabalhadores da saúde que ainda não foram vacinados contra a COVID-19 para escalonar uma nova etapa de vacinação para este grupo.
Transmissão de COVID-19 - Contato com dinheiro	Você pode pegar COVID-19 através de diversos elementos de contato, incluindo moedas e notas de dinheiro.

3.4.2.4. Escrever seções curtas

Seções curtas dão a sensação de que vamos entender de maneira mais fácil. Seções densas e longas são visualmente desagradáveis e dão a impressão de que a informação é difícil de entender. Para darmos títulos, uma seção curta também nos ajuda, pois, tem apenas um assunto em geral, o que não acontece com seções longas, pois são mais difíceis de resumir. Muitas vezes tratam de mais de um assunto. A Tabela 4 mostra um exemplo.

Tabela 3.4. Exemplo de seções curtas

Texto Comum	Texto com Linguagem Simples
<p>O programa tem uma linha telefônica Punjabi em 604-877-6163 na qual você pode deixar uma mensagem de correio de voz para alguém retornar sua chamada (as chamadas só são retornadas às sextas-feiras entre 10h00 e 12h00). Além disso, atendentes de reservas que falam chinês geralmente estão disponíveis pelo telefone 1-800-663-9203. Se não houver um disponível para atender sua chamada, você será questionado se alguém pode traduzir em seu nome ou para ligar de volta quando houver assistência de tradução disponível. Você também pode pedir a um membro da família ou amigo para ajudá-lo com a ligação.</p>	<p>O programa tem uma linha telefônica Punjabi em 604-877-6163 que você pode deixar uma mensagem de voz para alguém retornar sua chamada às sextas-feiras entre 10h00 e 12h00.</p> <p>Atendentes de reservas que falam chinês estão disponíveis pelo telefone 1-800-663-9203. Se não houver um disponível para atender sua chamada, será indicado ligar de volta quando houver tradução disponível.</p> <p>Você pode pedir a um membro da família ou amigo para ajudá-lo com a ligação.</p>

3.4.3. Escrita do Texto

A escrita do conteúdo é a terceira prática fundamental. A escolha cuidadosa das palavras, os verbos que devem dizer exatamente o que será feito, o cuidado com as abreviações fazem parte desta prática. Aqui, após termos nos preocupado com a estrutura, temos que nos preocupar com a forma de transmitir o conteúdo. As práticas parecem isoladas, mas não são. As diretrizes de escrita do texto na Linguagem Simples está direcionada para palavras, sentenças e parágrafos.

3.4.3.1. Palavras

A escolha das palavras na escrita do texto é algo muito importante. Podemos dizer que elas são o bloco mais básico do texto. Algumas características devem ser observadas quando estamos escolhendo palavras:

- **Concisão:** Escolha sempre o menor número de palavras possível para dizer a mesma coisa.
- **Precisão:** Escolha sempre a melhor palavra para transmitir uma ideia. Buscar sinônimos é sempre uma ótima prática.
- **Simplicidade:** Se tiver que escolher entre mais de uma palavra com o mesmo significado, escolha a mais simples, a que tem mais oportunidades de ser entendida pela sua audiência.
- **Domínio comum:** Escolha sempre as palavras de domínio da maior parte do seu público-alvo.

Outra prática que traz bons resultados é usar a voz ativa. A voz ativa é a voz verbal que indica que o sujeito da oração pratica a ação. E na Linguagem Simples sempre indicamos falar diretamente com quem é o responsável pela ação. Assim sendo ela será sempre a mais indicada.

Quanto aos verbos, use sempre as formas mais simples. O presente do indicativo é o tempo verbal mais recomendado. Verbos ocultos não devem ser utilizados. Em geral, aparecem convertidos em substantivos o que dificulta o entendimento do texto. Além disso, acaba obrigando que exista um verbo extra para completar o sentido da frase. Inca-se então que a conversão de verbos em substantivos seja minimizada o máximo possível.

Outro ponto importante é evitar o uso de jargões ou termos técnicos, legais ou estrangeiros quando existirem outras palavras com mesmo significado que possam substituí-los. Sempre é melhor usar palavras conhecidas da audiência. É importante também evitar ou reduzir ao máximo o uso de abreviações. Muitas vezes para poder entendê-las o leitor terá que repetidamente recorrer a outro local do texto ou da imagem onde esta está explicada. Neste ponto podemos falar também da boa prática de evitar apresentar definições. Em geral, elas são apresentadas no meio do texto, o que gera uma quebra de continuidade da ideia que está sendo apresentada.

O uso de palavras curtas também é recomendado. Se tiver que escolher entre duas conhecidas pela audiência e que tenham o mesmo significado, opte pela menor. Evite também ter que apresentar sua definição.

Além destas práticas citadas na estrutura das palavras, recomenda-se usar sempre o mesmo termo para se referir a um mesmo objeto. Neste caso podemos exemplificar o termo “patrão” e “empregador” sendo sinônimos, mas que para um público mais leigo pode levar a uma interpretação de que o primeiro tem relação com emprego informal e o segundo com emprego formal. A Tabela 5 nos mostra alguns exemplos.

Tabela 3.5. Alguns exemplos de uso de palavras

Texto Comum	Texto com Linguagem Simples
Os agendamentos podem ser feitos em diversos locais em todo o município.	Você pode fazer a marcação em qualquer local da cidade.
O relatório foi elaborado por João.	João escreveu o relatório.
Se o paciente sentir dor no peito, suor e falta de ar, deve ligar imediatamente para 190.	Se você sentir dor no peito, suor e falta de ar, ligue para 190.
Ao final será solicitado que o cliente dê seu feedback sobre o serviço.	Ao final pedimos a você que informe seu grau de satisfação com o serviço.

Traga sua carteira de identidade e/ou carteira de habilitação.	Traga um dos seguintes documentos: <ul style="list-style-type: none"> • Carteira de identidade • Carteira de habilitação
--	--

3.4.3.2. Sentenças

Escolha suas palavras e as organize com cuidado. Comece sempre com a ideia principal. As exceções devem ser tratadas depois. A ordem das palavras é muito importante. Coloque suas palavras com cuidado procurando ser sempre breve.

Não é ruim usar muitos pontos. É melhor escrever frases curtas com uma única ideia, do que longas com várias ideias nela. Isso vai acabar levando para outra boa prática de escrever sentenças também curtas.

A ordem natural das palavras de uma frase é sujeito-verbo-objeto. Busque sempre escrever desta forma. Sempre procure escrever frases completas. Neste caso de estrutura de frases, outro ponto importante são as negativas. Elas devem sempre ser evitadas. Estamos acostumados a pensar e falar positivamente. Negativas tendem a nos confundir. No caso do uso de duplas negativas e exceções isso fica ainda um pouco mais complexo para o entendimento. Quando escrevemos no negativo, colocamos um empecilho no caminho do entendimento do público-alvo. Quando você escreve uma frase contendo dois negativos, eles se cancelam, ou seja, na verdade, ela seria uma afirmativa. A melhor prática então é sempre usar a frase afirmativa, colocando a ideia principal no início e as exceções depois. A Tabela 6 nos mostra alguns exemplos.

Tabela 3.6. Alguns exemplos de sentenças

Texto Comum	Texto com Linguagem Simples
O treinamento, que contou com a participação de 60 funcionários e obteve índices de aprovação acima de 95%, foi um grande sucesso.	O treinamento foi bem-sucedido: 60 funcionários compareceram e deram índices de aprovação de mais de 95%.
Nenhum trabalho adicional é necessário na ausência de uma emergência.	O trabalho adicional só é necessário em caso de emergência.
Todas as pessoas, exceto aqueles com 18 anos ou mais, devem preencher o formulário.	Cada pessoa com menos de 18 anos deve preencher o formulário.
Ana encontrou o gerente da loja com o seu irmão.	Ana encontrou o seu irmão com o gerente da loja. OU Ana encontrou o gerente da loja com o irmão dele?

O recurso ao Tribunal de Primeira Instância, embora organizado de forma ineficaz, foi, no entanto, apresentado pelo escritório de advogados.	O escritório de advocacia apresentou seu requerimento ao Tribunal de Primeira Instância, apesar da organização ineficaz.
--	--

3.4.3.3. Parágrafos

Escreva parágrafos curtos que contenham somente uma ideia. Inicie sempre com uma sentença que define o tópico que vai ser detalhado. Uma frase de tópico pode fornecer uma transição de um parágrafo para outro. Uma palavra ou frase de transição diz claramente ao público-alvo se: (i) o parágrafo se expande no parágrafo seguinte, (ii) se contrasta com ele, (iii) se haverá uma direção completamente diferente. Outro ponto bastante importante é o título do parágrafo. Este deve conter um resumo do mesmo de forma que vá consumir tal informação já tenha um resumo da mesma pelo título. A Tabela 7 nos mostra alguns exemplos.

Tabela 3.7. Alguns exemplos de uso em parágrafos

Texto Comum	Texto com Linguagem Simples
Quando duas ou mais declarações são feitas em substituição e uma delas, se feita de forma independente, seria suficiente, a argumentação não é insuficiente pela insuficiência de uma ou mais das declarações alternativas.	Se uma das partes fez declarações alternativas, o pedido é suficiente se qualquer uma delas for suficiente.

3.4.4. Boas Práticas Gerais

Além de todas as práticas citadas anteriormente, algumas outras mais gerais também podem ajudar na melhoria do entendimento e na clareza da informação. Aqui temos algumas delas:

- O uso de exemplos pode ajudar seu público-alvo a entender melhor a informação.
- A criação de listas e tabelas pode organizar melhor um texto muito extenso e cheio de itens.
- A inclusão de uma ilustração pode ser mais útil do que uma extensa descrição textual.
- O uso de recursos como negrito e itálico podem ser interessantes para destacar coisas importantes.
- A redução do número de referências cruzadas pode ajudar o consumidor da informação a não se dispersar muito.

3.4.5. Boas Práticas acerca do Design

Na parte de design da informação outras práticas interessantes podem ser também utilizadas para melhoria do entendimento. A Figura 5 é um exemplo de como o design pode transformar a identidade visual de um elemento de informação. Aqui temos algumas destas boas práticas de design:

- Buscar sempre apresentar informações ordenadas e não muito densas. Em geral, quando não seguimos essa prática criamos uma reação negativa na mente dos usuários.
- Devemos dar bastante atenção ao design da informação. Ele é uma parte importante do desenvolvimento de um documento eficaz. Documentos agradáveis à vista são muito mais fáceis de entender do que os estilos mais tradicionais.
- Sempre que puder deve-se substituir blocos de texto por cabeçalhos, tabelas e listas para criar mais espaços em branco e deixar a apresentação mais leve ao olhar de quem a consome.
- Frases e seções curtas, como recomendado em práticas citadas aqui anteriormente, também dividirão um texto em partes visualmente melhor gerenciáveis.

CONTRATO DE PRESTAÇÃO DE SERVIÇOS

Pelo presente instrumento particular e na melhor forma de direito, as Partes:

(A) **PARTE A**, [nacionalidade], solteira, [profissão], portadora do documento de identidade RG n. [número do RG] [órgão emissor do RG], inscrita no CPF/MF sob o n. [número do CPF], residente e domiciliada na [Rua, Avenida, Praça...], n. [número], [Bairro], cidade de [Cidade], Estado de São Paulo, CEP [número do CEP] ("Parte A");

Sendo Parte A também denominada **CONTRATANTE**.

(B) **EMPRESA A**, sociedade simples limitada, com sede na [Rua, Avenida, Praça...], n. [número], [Bairro], cidade de [Cidade], Estado de São Paulo, [CEP], inscrita no CNPJ/MF sob o n. [número do CNPJ], neste ato representada por seus representantes legais ("Empresa A");

Sendo Empresa A também denominada **CONTRATADA**.

Sendo a **CONTRATANTE** e a **CONTRATADA** doravante denominadas em conjunto "Partes", e individualmente, "Parte".

RESOLVEM as Partes, de comum acordo, celebrar o presente Contrato de Prestação de Serviços (o "Contrato" ou "Instrumento"), nos termos da Lei n. 10.406 de 2002 ("Código Civil"), o qual será regido mediante as seguintes cláusulas e disposições:

CLÁUSULA 1 - OBJETO E ASPECTOS DOS SERVIÇOS

1.1. Objeto. O presente Instrumento tem por objeto a prestação de serviços de consultoria pela **CONTRATADA** à **CONTRATANTE**, os quais serão desempenhados com lealdade, zelo e eficiência (os "Serviços").

1.1.1. São serviços a serem prestados pela **CONTRATADA**: mapeamento de processos.

1.2. Alteração dos Serviços. Os Serviços indicados na **Cláusula 1.1** poderão ser alterados mediante comum acordo entre as Partes, por meio de instrumento escrito, devendo a remuneração específica envolvendo o conjunto de Serviços ser acordada previamente entre as Partes.



Figura 3.5. Exemplo de contrato com boas práticas de Design [Ribeiro, 2021].

3.5. Conclusão

Neste Capítulo apresentamos como a técnica da Linguagem Simples pode apoiar a construção de conteúdos de fácil entendimento. Um requisito necessário para a transparência das informações e seu uso para a transformação digital que vivenciamos. Destacamos os principais conceitos relacionados aos temas, demonstrando o inter relacionamento de cada tópico. No caso de órgãos públicos, por exemplo, a transformação digital promove ferramentas digitais para atuação eficiente do próprio

governo com participação de toda a sociedade. O foco na transparência é reforçado neste processo para que as pessoas possam usar serviços, dados e informações, ou mais precisamente, que todo esse conteúdo possa ser entendido de fato. A técnica da Linguagem Simples é usada na busca desse objetivo porque favorece o entendimento para uma efetiva transparência, e com isso impulsiona a transformação digital.

Foram apresentadas ainda as principais iniciativas no Brasil e no mundo no uso desta técnica de Linguagem Simples. Como pode ser visto hoje já contamos com diversas organizações que fazem coro na aplicação desta técnica sendo estas, centros de pesquisa, organizações internacionais independentes, universidades e outras, o que nos mostra a importância e magnitude deste movimento, chamado de Plain Language Movement, para garantir à sociedade o direito ao entendimento.

A partir do *Federal Plain Language Guidelines* apresentamos também as principais diretrizes para uma escrita mais simples. Foram mostradas práticas gerais para organização do texto como: definir o público-alvo adequado, usar bons títulos e escrever seções curtas de conteúdo. Seguindo a discussão para a escrita do texto propriamente dito, repassando orientações sobre as palavras, sentenças e parágrafos. Finalizando com boas práticas gerais sobre o design, o uso de exemplos e representações visuais. Este último, é um caso a parte que exploramos em pesquisas, mas que aqui fica como trabalhos futuros. Os recursos visuais naturalmente exigem menor esforço cognitivo à interpretação humana pelas suas formas. A Linguagem Simples, se propõe a melhorias não apenas no âmbito da escrita de texto, mas na estrutura e no design da informação. Contudo, no *Federal Plain Language Guidelines* há uma única diretriz que se refere ao uso de imagens para clareza informacional, indicando apenas o uso ilustrativo. Não há nenhuma menção a como se deve construir eficientemente esses recursos visuais. O *Federal Plain Language Guidelines*, poderia oferecer também um conjunto oficial de orientações para apresentação de informações e seu design em Linguagem Simples. Alguns estudos mais recentes já apontam como isso poderia ser feito.

Referências

- Appio, F. P., Frattini, F., Petruzzelli, A. M., & Neirotti, P. (2021). Digital transformation and innovation management: A synthesis of existing research and an agenda for future studies.
- M., Neirotti, P. (2021). Digital Transformation and Innovation Management: A Synthesis of Existing Research and an Agenda for Future Studies. *Journal of Prod Innov Manag*, 38(1), pp. 4-20.
- Barboza, E. M. F. (2010). A linguagem clara em conteúdos de websites governamentais para promover a acessibilidade a cidadãos com baixo nível de escolaridade. *Inc. Soc.*, Brasília, DF, v. 4 n. 1, pp. 52-66.
- Brasil. Mais de 70% dos serviços públicos brasileiros já são digitais. Agência Brasil. (2021). Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2021-07/mais-de-70-dos-servicos-publicos-brasileiros-ja-sao-digitais>. Acesso em: 13 agosto de 2021.

- Brasil. Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC). Estratégia Brasileira de Transformação Digital: E-digital. (2018). Disponível em: <http://www.mctic.gov.br/mctic/export/sites/institucional/estrategiadigital.pdf>. Acesso em: 13 agosto 2021.
- Boscarioli, C., Araujo, R.M, Maciel, R.S. (2017). I GranDSIBR: Grand Research Challenges in Information Systems in Brazil 2016-2026. Special Committee on Information Systems. Brazilian Computer Society (SBC). ISBN: [78-85-7669-384-0]. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/book/28>
- Barcellos, R. (2017). Avaliação da Qualidade e Interpretabilidade de Visualizações de Dados. (Dissertação de Mestrado, Universidade Federal Fluminense) Niterói. pp. 87.
- Cappelli, C., Leite, J. C. S. P. (2008). Transparência de processos organizacionais. II Simpósio Internacional de Transparência nos Negócios, Universidade Federal Fluminense, LATEC, Niterói, RJ, Brasil.
- Cappelli, (2009) Uma abordagem para transparência em processos organizacionais utilizando aspectos. Tese de Doutorado. Disponível em: <http://www-di.inf.puc-rio.br/~julio/tese-cappelli.pdf>
- Collin, J., Halen, M., Helenius, M., Hiekkanen, K., Itälä, T., & Korhonen, J. (2014). IT Leadership in Finnish Organizations and Digital Transformation. Research Report. Department of Computer Science, Emeritus Mäntylä Martti.
- Dias, D. T. M. A. R. (2019). Fatores chaves de sucesso para transformação digital. Dissertação de Mestrado. Fundação Getúlio Vargas, Escola Brasileira de Administração Pública e de Empresas, Centro de Formação Acadêmica e Pesquisa.
- Fischer, H.; Mont'alvão, C.; Rodrigues, E. S.; Engelke, A. (2019), Compreensibilidade em textos de e-gov: uma análise exploratória da escrita do INSS. In: 9º Congresso Internacional de Design da Informação, p. 303-313. Disponível em: <https://www.proceedings.blucher.com.br/article-details/33627>
- INAF, B. Indicador de Alfabetismo Funcional. (2018). Disponível em: https://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf. Acesso em: 12 ago. 2021.
- Govinfo. Public Law 111 - 274 - Plain Writing Act of 2010 (2010). Disponível em: <https://www.govinfo.gov/app/details/PLAW-111publ274>>. Acesso em: 07 ago. 2021.
- Henriette, E. Feki, M., Boughzala, I. (2015); The Shape Of Digital Transformation: A Systematic Literature Review; MCIS Proceedings, Paper 10, pp. 1-19.
- Kane, C., Palmer, D., Phillips, A. N., Kiron, D., Buckley, N. (2017) Achieving digital maturity. MIT Sloan Management Review, v. 59, n. 1.
- Layne, K., Lee, J. (2001). Developing fully functional e-government: A four stage model. Government information quarterly, Elsevier, v. 18, n. 2, p. 122–136.
- Laudon, K. C., Laudon, K. P. (2016). Managing information system: Managing the digital firm (14th ed.). England: Pearson.

- McNutt, K. (2014). Public engagement in the Web 2.0 era: Social collaborative technologies in a public sector context. *Canadian Public Administration*, 57(1), 49-70.
- Macalintal, M., Chepkasova, E. (2017). Transformation in the Era of Digitization: A study of organizations implementing digital transformation projects with integrated project management and change management. Dissertação de Mestrado. Umeå School of Business and Economics.
- Morakanyane, R., Grase, A. A., O'Reilly, P. (2017). Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature. In: Bled eConference, p. 21.
- Oliveira, R.; Cappelli, C.; Oliveira, J., (2021). Diretrizes para o Design de Visualização da Informação: Estendendo a Linguagem Cidadã. In: Workshop de Computação Aplicada em Governo Eletrônico (WCGE), p. 259-266. ISSN 2763-8723. Disponível em: <https://sol.sbc.org.br/index.php/wcge/article/view/15994>
- Park, S., Gil-Garcia, J. R. (2017). Understanding transparency and accountability in open government ecosystems: The case of health data visualizations in a state government. In Proceedings of the 18th Annual International Conference on Digital Government Research. pp. 39-47.
- Resnick, M., (2002). Rethinking Learning in the Digital Age.
- Ribeiro, A. Visual Law | Contrato. Behance, (2021). Disponível em: https://www.behance.net/gallery/105020447/Visual-Law-Contrato?tracking_source=search_projects_recommended%7Ccontrato. Acesso em: 30 de agosto de 2021.
- Rogers, D. L. (2016). The Digital Transformation Playbook: Rethink your Business for the Digital Age. New York: Columbia University Press.
- Sandberg, J., Mathiassen, L., Napier, N. (2014). Digital options theory for IT capability investment. *Journal of the Association for Information Systems*, v. 15, n. 7, p. 1.
- Shumskaya, D., (2019). Linguagem Clara: uma questão de cidadania corporativa para as instituições financeiras? (Doctoral dissertation, Instituto Politécnico de Lisboa, Escola Superior de Comunicação Social).
- Song, C., Lee, J. (2016). Citizens' use of social media in government, perceived transparency, and trust in government. *Public Performance & Management Review*, 39(2), 430-453.
- Tadeu, H. F. B., Duarte, A. L. C. M., Chede, C. T. (2018). Transformação Digital: perspectiva brasileira e busca da maturidade digital. *Revista DOM. Fundação Dom Cabral. Nova Lima, DOM: v.11, n.35*, p. 32-37, maio/ago. 2018. Acesso em: <https://acervo.ci.fdc.org.br/AcervoDigital/Artigos%20FDC/Artigos%20DOM%2035/Transforma%C3%A7%C3%A3o%20Digital%20perspectiva%20brasileira%20e%20busca%20da%20maturidade%20digital.pdf>.



Claudia Cappelli

Professora do curso de Ciência de Computação da UERJ. Colaboradora do programa de Pós Graduação em Sistemas de Informação da UFRJ. Doutora em Ciências - Informática pela PUC-Rio (2009). Mestre em Sistemas de Informação pela Universidade Federal do Rio de Janeiro (2000). Graduada em Informática pela Universidade do Estado do Rio de Janeiro (1985). Realizou estágio Pós-Doutoral junto ao Programa de Pós-Graduação em Informática da Unirio (2010) e na UFRJ (2020). Pesquisadora Jovem Cientista Nosso Estado pela FAPERJ. Fundadora do Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT-DD). Pesquisadora em Literacia Digital do IBCIH (Instituto Brasileiro de Cidades Inteligentes e Humanas). Pesquisadora em Linguagem Cidadã no LincLab (Laboratório Interdisciplinar de Linguagem Clara). Membro do Comitê Gestor do programa Meninas Digitais da SBC. Mentora da Fábrica de Startups. Pesquisadora em Inovação no INEI (Instituto Nacional de Empreendedorismo e Inovação). Foi Professora do Departamento de Informática Aplicada da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) (2005 - 2018).

<http://lattes.cnpq.br/4930762936357558>



Vanessa Nunes

Doutora em Engenharia de Sistemas e Computação pela COPPE/UFRJ desde 2014 e mestre em Sistemas de Informação pelo NCE/UFRJ desde 2007, tendo concluído a graduação em Ciência da Computação pela UFRJ em 2004. Pós-doutora pelo Departamento de Ciência da Computação da Universidade de Brasília (UnB) em pesquisas relacionadas aos temas de computação transparente, arquitetura corporativa, modelos de referência para avaliação e evolução de cidades inteligentes e humanas e sistemas multi-agentes. Atua como Professora Substituta no Departamento de Ciência da Computação da Universidade de Brasília. Membro de pesquisa da Rede Brasileira de Cidades Inteligentes e Humanas (RBCIH - www.redebrasileira.org) Participa também na organização da SE7Ti - Tecnologia e Inovação.

<http://lattes.cnpq.br/2043415661294559>



Rodrigo Oliveira

Mestrando em Informática, linha de pesquisa em Sistemas de Informação pela Universidade Federal do Rio de Janeiro, UFRJ (2019) possui Bacharel em Sistemas de Informação pela Universidade Federal do Estado do Rio de Janeiro, UNIRIO (2018) e Graduado em Design Gráfico pela Universidade do Grande Rio, UNIGRANRIO (2013). Pesquisador em Linguagem Cidadã no LincLab (Laboratório Interdisciplinar de Linguagem Clara). Analista de Geotecnologia na Imagem Geosistemas (img.com.br). Carreira desenvolvida na área de Design e Tecnologia, com experiência em comunicação visual, design de interação e desenvolvimento de software. Com interesses na área de Visualização de Dados, UX Design, Linguagem Simples e Gestão de Processos de Negócios, com ênfase em transparência da informação e automação de soluções em TI.

<http://lattes.cnpq.br/1091110605525620>

Capítulo

4

Técnicas e práticas de jornalismo de dados para aquisição e gerenciamento de dados em MySQL aplicadas ao domínio da violência contra a mulher

Luciana Sá Brito, Alayne Duarte Amorim, André Viana Tardelli, Angélica Fonseca da Silva Dias, Juliana Baptista dos Santos França, Adriana Santarosa Vivacqua

Abstract

This short course aims to provide a panoramic view of the fundamental steps of acquiring and managing data practice in data journalism. It will be conducted using MySQL databases resources and query language basic structures. Through the introduction of anonymization, cleansing, and transformation techniques, participants will be encouraged to practice these techniques in order to increase students' data literacy. A real case of data acquisition and management of violence against women will be applied to foment flashes of inspiration to this short course in order to amplify students' critical view and empathy about data journalism's social issues.

Resumo

Este capítulo tem como objetivo desenvolver uma visão panorâmica das etapas fundamentais de aquisição e gerenciamento de dados praticadas no âmbito do jornalismo de dados com o uso do MySQL. Nele, serão introduzidos conceitos e práticas associados às técnicas de anonimização, limpeza e transformação de dados, a fim de estimular a literacia de dados do leitor. Um caso real de aquisição e gerenciamento de dados de violência contra a mulher será utilizado como objeto de aplicação das técnicas e conceitos apresentados. Espera-se que ao final do capítulo, o leitor tenha sua visão crítica e empática ampliada, acerca de temáticas sociais aderentes ao jornalismo de dados.

4.1. Introdução e Contextualização do Problema

Uma em cada 3 mulheres no mundo possui histórico de violência física e/ou sexual [World Health Organization *et al.* 2020]. Em tempos de emergência, a violência contra as mulheres continua ameaçando seriamente a saúde da população, especialmente das mulheres, sendo que o tipo mais recorrente de violência é aquela realizada pelo seu próprio parceiro íntimo [World Health Organization *et al.* 2020].

Segundo a Organização Mundial de Saúde (OMS), violência é o uso intencional de poder ou força física, em forma de ameaça ou indo às vias de fato, contra si mesmo, outra pessoa, ou contra um grupo ou comunidade, que pode resultar em ferimentos, morte, dano psicológico, desenvolvimento deficiente ou privação [World Health Organization *et al.* 2020]. Violência contra a mulher é, segundo Burelomova *et al.* (2018), o mau uso do poder pelo parceiro íntimo (homem ou mulher), que resulta em perda de dignidade, controle e segurança, bem como sentimento de impotência e aprisionamento experimentado pela mulher que é vítima direta de problemas físicos, psicológicos, abuso econômico, sexual, verbal e/ou espiritual contínuos ou repetidos. Violência contra a mulher também inclui ameaçar ou forçar mulheres o testemunho de violência por parte de seus maridos, parceiros, ex-maridos ou ex-sócios, contra seus filhos, parentes, amigos, animais de estimação e/ou bens queridos.

A Lei nº 11.340 de 2006 [Brasil 2006] nomeada Lei Maria da Penha recebeu esse nome em razão da vítima Maria da Penha Maia Fernandes (Figura 4.1) e tem como objetivo principal estipular punição adequada e coibir atos de violência doméstica e familiar contra a mulher. Segundo a referida lei, as formas de violência são classificadas como: violência física, violência psicológica, violência sexual, violência patrimonial e violência moral.



Figura 4.1. Maria da Penha, a ativista pelos direitos das mulheres que dá nome à Lei 11.340/06, ficou paraplégica por conta de agressões realizadas por seu ex marido

Fonte:
[https://commons.wikimedia.org/wiki/File:Maria_da_Penha_em_novembro_de_2018_\(cropped_2\).jpg](https://commons.wikimedia.org/wiki/File:Maria_da_Penha_em_novembro_de_2018_(cropped_2).jpg)

O artigo 38 da Lei Maria da Penha prevê a criação de um Sistema Nacional de Dados e Estatísticas sobre a Violência Doméstica e Familiar contra a Mulher, que deveria ter sido implementado nos quatro anos seguintes à sua publicação pela Secretaria Especial de Políticas para as Mulheres (SPM) em articulação com outros Ministérios e órgãos da Administração Pública. Contudo, após cerca de 15 anos da publicação da Lei, ainda não é possível identificar um sistema nacional de dados sobre violência doméstica.

A ausência de uma base de dados nacional detalhada sobre a violência contra a mulher prejudica o reconhecimento da gravidade e alcance da violência doméstica no Brasil, favorecendo a manutenção das graves consequências para a saúde e segurança das mulheres do nosso país [Alves, Dumaresq e Silva 2016], problema que é agravado nas regiões do país onde há maior vulnerabilidade social [Borburema *et al.* 2017].

Este capítulo foi concebido a partir da compreensão de que é urgente aperfeiçoar as fases de coleta e tratamento de dados de violência doméstica, de forma que os resultados de sua análise reflitam essa realidade social, sendo capazes de orientar ações e políticas públicas capazes de garantir qualidade de vida e segurança às mulheres brasileiras.

4.1.1. Dados sobre Violência Contra a Mulher no Brasil

Em 2016, o Núcleo de Estudos e Pesquisas da Consultoria Legislativa do Senado nacional apontou lacunas existentes nas políticas de enfrentamento da violência contra a mulher, apontando a relevância de um banco de dados unificado sobre o tema [Alves, Dumaresq e Silva 2016]. O estudo indicou a existência de bases de dados pulverizadas, revelando a forma como os órgãos utilizam essas bases parciais para a geração de políticas através das análises desses dados.

O Observatório da Mulher Contra a Violência — órgão instituído pelo Senado Federal em 2016 para funcionar em conjunto com o Instituto DataSenado com a função de reunir e sistematizar as estatísticas oficiais sobre a violência contra a mulher — inaugurou, em março de 2019 o recurso Painel de Violência Contra Mulheres¹, que relaciona dados de diversas fontes para oferecer um panorama da violência contra a mulher no Brasil. O painel é carregado com dados coletados através do Ministério da Saúde e Conselho Nacional de Justiça, mais especificamente das fontes: Sistema de Mortalidade do Ministério da Saúde, Sistema de Informação de Agravos de Notificação do Ministério da Saúde, Pesquisa Nacional por Amostra de Domicílios do IBGE, Secretarias de Segurança Pública estaduais, Conselho Nacional de Justiça e Disque 180 da Secretaria de Políticas para Mulheres.

A navegação pelo Painel de Violência contra Mulheres contempla a visualização de dados de 2011 a 2018 através da aplicação de filtros por ocorrência de: homicídios de mulheres, notificações de saúde, boletins de ocorrência e ações do poder judiciário. Adicionalmente, podem ser aplicados filtros como ano, UF, raça, escolaridade entre outros. Contudo, ao realizar testes para a geração de visualizações de dados através do site, foram constatadas algumas inconsistências como a mostrada na Figura 4.2, que aponta a inexistência de boletins de ocorrência de violência contra mulheres no ano de 2017 — fonte de questionamento quanto a sua discrepância frente à realidade.

¹ <https://bit.ly/3vst86w>. Acesso em 27/05/2021.

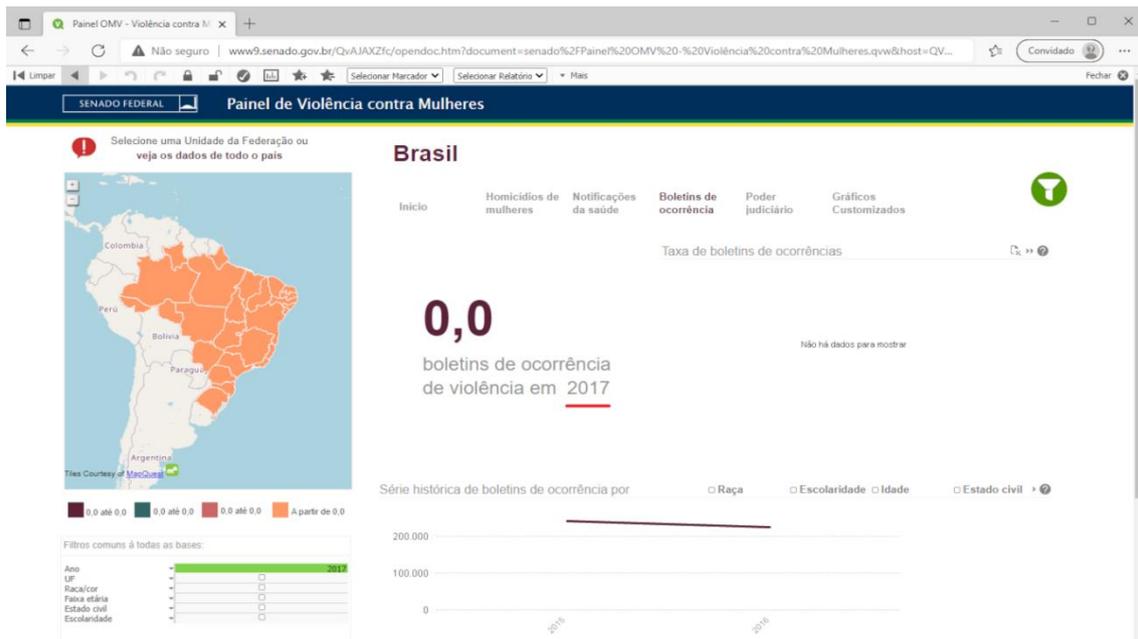


Figura 4.2. Painel Violência contra Mulher do Senado Federal. Aplicação de filtro por ano para visualizar boletins de ocorrência feitos em 2017.

O fato de não haver registros de boletins de ocorrência de violência doméstica em 2017 para todo o território nacional corrobora com a inquietação sobre a urgência da reunião dos dados como indicado na Lei Maria da Penha, porque reflete a percepção de que as bases de dados nacionais sobre violência contra a mulher realmente não estão unificadas.

Reconhecendo a pulverização das bases de dados sobre violência doméstica no Brasil, o Grupo de Pesquisa em Literacia de Dados do Programa de Pós graduação em Informática da Universidade Federal do Rio de Janeiro (PPGI - UFRJ) empreendeu um projeto com o objetivo de localizar um *dataset* referente ao registro de atendimentos a mulheres vítimas de violência doméstica e criar uma base de dados consolidada para a realização de aplicações de Ciência de Dados.

A partir dessa busca, foi realizado contato com um Centro Especializado de Atendimento à Mulher (CEAM) de um município do Estado do Rio de Janeiro, que disponibilizou um conjunto de dados para estudo. A escolha do município se deu por ser uma região historicamente marcada pela vulnerabilidade social e pela possibilidade de acesso às informações através do centro especializado.

A pesquisa realizada sobre esse conjunto de dados contribuiu para o estudo das técnicas, conceitos e práticas de jornalismo de dados abordadas neste capítulo. Ela também contribuiu para a devolução de uma base de dados consistente para o CEAM, possibilitando a visibilidade dos mais de 150 casos de violência doméstica registrados somente durante os anos iniciais da pandemia do Coronavírus (Figura 4.3), e possibilitando que outras organizações, pesquisadores e gestores produzam ou utilizem modelos para automatização da visualização dos dados para geração de conhecimento e publicações atualizadas sobre os dados de violência doméstica, fundamentando a intensificação de ações de combate a este tipo de crime.

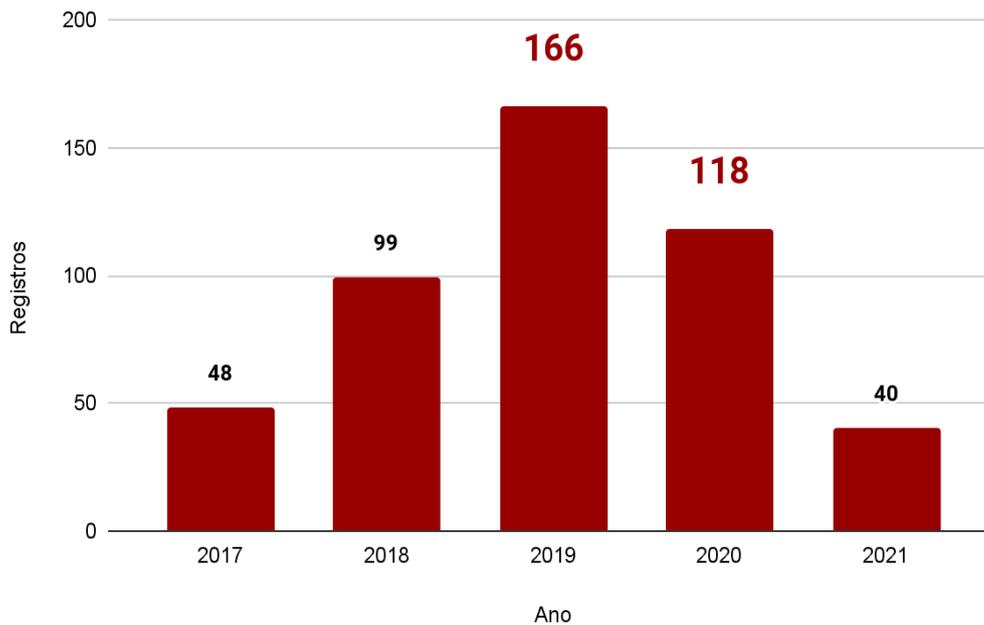


Figura 4.3. Gráfico com número de registros de violência entre os anos 2017 e 2021, com base nos dados do CEAM. A baixa quantidade de registros nos anos de 2017 e 2021 revela o início da contagem e contagem ainda em andamento.

A partir do encontro do Grupo de Literacia de dados do PPGI-UFRJ com o tema de estudo e com o conjunto de dados do CEAM, identificou-se que o Jornalismo de Dados seria uma abordagem interessante para lidar com todo o processo de descoberta de informação na base de dados. Na próxima seção, é dada uma visão geral sobre o Jornalismo de Dados.

4.1.2. Jornalismo de Dados

Compreender o jornalismo de dados pressupõe a compreensão de duas práticas jornalísticas anteriores: o jornalismo em seu sentido tradicional e o jornalismo investigativo. O jornalismo é a atividade profissional que consiste em lidar com notícias, dados factuais e divulgação de informações. É a prática de coletar, redigir, editar e publicar informações sobre eventos atuais, sendo fundamentalmente uma atividade de comunicação [Ferreira 2012]. Já o jornalismo investigativo é todo trabalho de jornalista que não se limita ao release, ao declaratório, mas vai ao cruzamento de dados e ao destrinchamento de documentos, à escuta dos personagens envolvidos e à busca de informações estatísticas [ABRAJI 2012].

O jornalismo de dados é uma evolução do jornalismo investigativo [Appelgren e Nygren 2014] que pode se apresentar como uma mistura entre o jornalismo tradicional, o jornalismo investigativo e um jornalismo que usa bases de dados de grandes volumes (*big data*), com o auxílio de mecanismos eletrônicos para a análise e o processamento de dados, a fim de encontrar e comunicar verdades sociais importantes. O que torna o jornalismo de dados diferente dos demais, é a capacidade de combinar o “faro” tradicional para notícias com a capacidade de contar histórias convincentes por meio de dados [Bounegru *et al.* 2012].

Os produtos do jornalismo de dados são, entre outros:

- conjuntos de dados abertos (ou *datasets*);
- infográficos;
- *apps* de notícias;
- *blogs* de dados;
- *sites* de dados;
- visualizações interativas.

Alguns exemplos de visualizações interativas são:

- Projeto “Sobreviventes”, que investiga a violência de gênero na Colômbia - <http://especiales.datasketch.co/sobrevivientes/index.html>;
- Projeto “Lost Mothers”, que dá visibilidade às altas taxas de mortalidade materna nos Estados Unidos - <https://www.propublica.org/article/lost-mothers-maternal-health-died-childbirth-pregnancy>;
- Projeto News 18 sobre a falta de reconhecimento do trabalho realizado pelas camponesas na Índia - <https://www.news18.com/news/immersive/women-farmers-of-india.html>.

Neste capítulo não foi utilizado um *dataset* caracterizado como *big data*, porque a finalidade pedagógica de interesse é prover conhecimentos básicos de gerenciamento de dados em MySQL. Conhecimentos esses, que uma vez aprendidos e praticados com um *dataset* de pequenas proporções, poderá ser aplicado em conjuntos de dados mais volumosos.

4.1.3. Tipos de Acesso, Licença e Formato

Os dados que o jornalista utilizará na sua investigação poderão ser fornecidos por entidades públicas por meio de licenças mais restritivas (fechadas) ou mais permissivas (abertas). Há várias questões jurídicas presentes nessa questão, que variam em função da licença adotada por cada entidade.

As licenças abertas, em geral, permitem o livre uso e reuso dos dados [Villars *et al.* 2011]. Como exemplo de disponibilização de dados abertos, temos o *site* atual do Ministério da Saúde², que informa em seu rodapé: “Todo o conteúdo deste site está publicado sob a licença Creative Commons Atribuição - SemDerivações 3.0 Não Adaptada”. Já como exemplo de disponibilização de dados fechados, temos o *site* antigo do Ministério da Saúde³, que informa em seu rodapé “Copyright © Ministério da Saúde. Todos os direitos reservados 2013/2021”.

Os governos em geral mantêm plataformas com dados em acesso aberto (ou livre), que são os dados que podem ser acessados gratuitamente mesmo que a liberdade de uso desses dados possa variar segundo sua licença. Alguns *sites* de dados de livre acesso governamentais são:

² <https://www.gov.br/saude/pt-br>. Disponibilizado em agosto de 2021.

³ <https://antigo.saude.gov.br/>. Disponibilizado em agosto de 2021.

- Governo da Colômbia: <https://www.datos.gov.co/>;
- Governo da Argentina: <https://www.datos.gob.ar/>;
- Governo do Brasil: <https://dados.gov.br/>;
- Governo do México: <https://datos.gob.mx/>;
- Governo dos EUA: <https://catalog.data.gov/dataset>;
- Governo da Rússia: <https://data.gov.ru/?language=en>.

É importante não confundir acesso aberto com licença aberta. Dados de acesso aberto são os dados que se pode acessar gratuitamente, enquanto os dados de licença aberta, mais conhecidos como os “dados abertos” são aqueles dados que podem ser consumidos livremente e redistribuídos em sua versão original ou modificada.

A Open Knowledge Foundation define que “Dados e conteúdo abertos podem ser usados, modificados e compartilhados livremente por qualquer pessoa para qualquer propósito” sendo relevante observar a licença desses dados, pois requisitos de preservação de procedência e abertura podem existir [Open Knowledge Foundation 2021]. A partir da tomada de consciência do valor social dos dados, especialistas em dados abertos têm discutido sobre a abertura de dados do setor privado para ações que beneficiem o interesse público, fomentando a colaboração por meio da troca de dados entre os setores (público/privado), o chamado *Data Collaboratives* [GovLab, 2021].

Além do acesso e da licença, é necessário observar o formato no qual os dados estão apresentados. Formatos abertos de arquivos, em geral, podem ser processados por diversos programas de computador e incluem formatos como .jpg, .mp3, .pdf, bem como conjuntos de formatos, tais como os *OpenDocuments* (.odt, .ods, etc.) e os *Office Open XML* (.docx, .pptx, .xlsx, etc.). Os formatos fechados são protegidos por patentes e costumam ser lidos apenas pelos programas oficiais das empresas que detêm essas patentes. Como exemplos de formatos fechados, temos os .xls, .doc, .ppt, .bmp e outros. O formato de arquivo mais adequado para a publicação de dados abertos é o .csv (*Comma Separated Values* - Valores Separados por Vírgulas), que apresenta os valores organizados como mostra a Figura 4.4.

linha	atributos
1	ficha_ano,"ano","usuaria","contato","tecnica_de_ref","idade","bairro","municipio","origem","tipos_de_vd","encam","dia","religiao","grau_escolaridade","cor","com_renda","filhos","rel_agressor"
2	01*2020,"2020","India Viana Pimentel","5(3527)586-74-47","Marina Souza","66","Arcos","PORTO","ONG","n","np","1/2/2020","np","ensino mÃ©dio","negra","nÃ£o","1","Ex marido"
3	02*2020,"2020","Sofia Boto da Costa","791(1715)084-08-41","Marina Barbosa","59","Moriadeira","PORTO","ONG","C/E/D/R","T/C","1/2/2020","catÃ©lica","ensino fundamental","negra","sim","3","outros"
4	03*2020,"2020","Jane GourjÃ£o SalomÃ£o","34(6241)194-57-42","Marina Barbosa","25","arcos","PORTO","AvÃ¡","C/E/D","T/A/D/Vara","1/3/2020","luterana","alfabetizaÃ§Ã£o","branca","sim","8","marido"
5	04*2020,"2020","Bianca ValadÃ£o Coimbra","60(3238)132-07-44","Joana","28","np","PORTO","Vizinha","C/E/D/re","DEL/D","1/4/2020","catÃ©lica","ensino mÃ©dio","amarela","nÃ£o","1","irmÃ£"
6	05*2020,"2020","Ester Escobar Grande","9(717)298-74-52","Marina Barbosa","32","Rechousa","PORTO","APAV","C/E/D/RC","terap","1/4/2020","calvinista","ensino mÃ©dio","negra","sim","2","marido"
7	06*2020,"2020","Ludmila RebouÃ§as Bugalho","187(2931)545-52-99","Marina Barbosa","41","CastelÃ¶es","PORTO","ONG","C/E/D/R","terap","1/5/2020","np","np","branca","sim","3","mÃ£e"
8	07*2020,"2020","Ellen Leme Mirandela","83(01)200-94-51","Bruna","22","Agro Velho","PORTO","Conselho tutelar","C/E/D/R","np","1/8/2020","budista/xintoista","ensino mÃ©dio","amarela","sim","3","marido"
9	08*2020,"2020","Roberta Freiria Castelhana","7(97)803-15-86","Marina Barbosa","64","Moninhas","PORTO","ONG","C/E/D/re","AM/MÃ©dico de famÃ©lia","1/8/2020","budista","alfabetizaÃ§Ã£o","amarela","sim","1"
10	09*2020,"2020","Neusa Pacheco Matias","1(055)755-97-59","Marina Barbosa","38","CoimbrÃ¶es","PORTO","Conselho tutelar","np","terap","1/10/2020","atÃ©lia","ensino fundamental","mestiÃ§a","sim","2","hora"
11	10*2020,"2020","Maira Leiria Quinzeiro","959(29)929-06-15","Marina Barbosa","52","Bairro de Aldoar","PORTO","Igreja","C","Terap","1/12/2020","catÃ©lica","alfabetizaÃ§Ã£o","amarela","sim","1","np"
12	11*2020,"2020","Nara Botelho RosÃ¡rio","76(462)094-59-21","Joana","17","Moninhas","PORTO","Amiga","C/E/RC","A/T","1/15/2020","np","alfabetizaÃ§Ã£o","negra","np","7","marido"
13	12*2020,"2020","MelÃ¢nia JordÃ£o","913(2603)126-74-12","Marina Barbosa","np","Arcos","PORTO","APAV","C/E/D","terap","1/16/2020","catÃ©lica","alfabetizaÃ§Ã£o","mestiÃ§a","sim","2","marido"
14	13*2020,"2020","Larissa Chaves AragÃ£o","8(5504)091-91-28","Joana","51","Leandro","PORTO","FuncionÃ¡rio do APAV","C/S/D/rc/e","ter","1/22/2020","matriz afro","ensino mÃ©dio","negra","sim","4","companhei"

Figura 4.4. Tabela de dados em formato .csv exibidos no Microsoft Excel. © Microsoft Corporation.

4.2. Coleta dos Dados

O *dataset* é uma parte essencial de um projeto de descoberta de informações em uma base de dados, e encontrar um conjunto de dados que atenda às necessidades de pesquisa é boa parte do trabalho. De uma forma geral, os dados podem ser obtidos em repositórios disponibilizados para acesso e uso público (veja exemplo no Vídeo 1⁴) ou por iniciativas próprias de coleta, através de metodologia e recursos específicos para esse fim, como: pesquisas, questionários, entrevistas e outras formas que permitam reunir informações úteis e organizadas no formato desejável.

Ferramentas como o Kaggle⁵ facilitam a aquisição de dados através dos *datasets* disponibilizados na sua plataforma, enquanto que ferramentas como o Google Data Search⁶ facilita a aquisição de dados que estão disponibilizados em diferentes plataformas. O Kaggle é um dos ambientes mais conhecidos para competições de Ciência de Dados, armazenando e disponibilizando dados sobre assuntos diversos. O Google Data Search é uma ferramenta de busca que localiza um *dataset* onde quer que ele esteja hospedado. Seu funcionamento é semelhante ao mecanismo de busca do Google Scholar.

Contudo, apesar de podermos contar com diversos meios eletrônicos de disponibilização de conjuntos de dados, é possível não encontrar um conjunto que atenda às expectativas, ou ainda, não encontrar um que esteja disponibilizado para acesso público. Essa última foi a situação que o grupo de pesquisa se deparou quando iniciou a investigação sobre violência doméstica. Foram realizadas buscas por bases de dados abertos governamentais relacionados ao tema de pesquisa, considerando o período dos últimos 10 anos que, em teoria, corresponde ao tempo máximo indicado na Lei Maria da Penha para a implantação do Sistema Nacional de Dados e Estatísticas sobre a Violência Doméstica e Familiar contra a Mulher.

A partir da constatação da dificuldade de encontrar uma base de dados unificada e com acesso público sobre o tema, este estudo empenhou esforços em localizar os dados referentes aos registros de violência doméstica contra mulher em um dos municípios do Estado do Rio de Janeiro, conforme descrito na seção 4.1.1. A primeira forma de busca aconteceu através de contato telefônico com a Secretaria de Assistência Social do município, que direcionou os pesquisadores para o Centro Especializado de Atendimento à Mulher (CEAM).

Em entrevista com os especialistas do CEAM, foi identificado que os registros de atendimentos a mulheres vítimas de violência doméstica acontecem em formulários impressos em papel, sendo posteriormente digitados em planilhas eletrônicas. Constatou-se que a geração dessas planilhas faz parte de uma iniciativa independente do CEAM, uma vez que não há uma indicação de que esses dados alimentam uma base integrada no município ou em outras instâncias governamentais. Contudo, os especialistas relataram que remetem relatórios periódicos aos gestores do município a partir desses dados planilhados.

O *dataset* do CEAM contém os registros dos atendimentos realizados na unidade às mulheres vítimas de violência doméstica e está dividido em 5 planilhas eletrônicas que refletem as características da população atendida, bem como dados contextuais sobre a

⁴ O Vídeo 1 pode ser acessado em <https://youtu.be/EvYvKUeXqis>.

⁵ <https://www.kaggle.com/>.

⁶ <https://datasetsearch.research.google.com/>.

violência sofrida no período de outubro de 2017 a março de 2021, época da cessão do conjunto de dados pela diretoria do CEAM. Devido à natureza sensível dos dados, o CEAM não os disponibiliza como dados abertos nem autoriza seu acesso ao público. Por esse motivo, foi requerido aos integrantes do grupo de pesquisa um termo de uso de dados e anonimato dos indivíduos listados nos registros. Esta pesquisa se reserva na obrigação de não publicar a base de dados e seu conteúdo.

Para a realização das atividades práticas descritas neste capítulo será utilizado um *dataset* fictício, chamado *mulher_2020*⁷, criado a partir do conjunto de dados da CEAM. Na próxima seção, será abordado o processo de criação de uma base de dados utilizando o *dataset* *mulher_2020*, à semelhança da criação da base de dados gerada com o *dataset* CEAM através dos processos de carregamento, anonimização, limpeza e transformação.

4.3. Criação da Base de Dados

Um banco de dados é uma coleção organizada de dados e informações estruturadas, geralmente armazenadas eletronicamente em um sistema de computador [Oracle 2021]. Os bancos de dados têm como objetivo dar suporte aos sistemas de informação e geralmente são controlados por um Sistema de Gerenciamento de Banco de Dados (SGBD), que gerencia a manipulação dos dados, dando suporte a consultas e permitindo criar cópias de segurança, entre outras diversas tarefas [Bazzi 2013].

O MySQL é um SGBD da categoria dos *softwares livres*, pertencente à Oracle Corporation, que usa como interface a Linguagem de Consulta Estruturada (SQL). Algumas grandes corporações utilizam o MySQL para lidar com grandes volumes de dados, como o Facebook, a Google e a Adobe [Oracle Corporation and/or Its Affiliates 2021].

Este capítulo utiliza o MySQL como suporte para a criação de uma base de dados utilizando o conjunto de dados *mulher_2020*. Através das consultas SQL serão realizadas as fases de anonimização, limpeza e transformação dos dados a fim de gerar uma base de dados consolidada para a realização de análises que permitam mostrar para a sociedade características relevantes da violência contra mulheres em uma determinada localidade.

Na próxima seção, será dado início às atividades práticas de jornalismo de dados, iniciando com a ativação do banco de dados MySQL.

4.3.1. Banco de Dados MySQL em Ação

Para iniciar a criação da base de dados para análise dos dados de violência contra a mulher, será configurado um banco de dados MySQL. Para realizar essa configuração, primeiramente é necessário ativar o servidor MySQL e depois determinar a escolha de uma *interface* gráfica para o gerenciamento dos dados da base.

Para iniciar o servidor de banco de dados, será utilizada a ferramenta XAMPP, que é uma distribuição Apache e constitui-se como um pacote que contém os principais servidores de código aberto existentes, entre eles o MySQL. É possível obter a ferramenta

⁷ O *dataset* *mulher_2020* pode ser acessado no repositório <https://github.com/Lu-Brito/Escola-Regional-de-Sistemas-de-Infirma-o-2021---UNIRIO>.

XAMPP através do *site XAMPP Installers and Downloads for Apache Friends*⁸. Após baixá-lo, você deverá abrir o arquivo executável e realizar o passo a passo de instalação que você encontrará no Vídeo 2⁹.

Depois de instalado o servidor, deverá ser instalado adicionalmente o *software* que permitirá a criação da base de dados e realização das consultas SQL. Para esta prática, será utilizado o MySQL Workbench, que é a ferramenta de *interface* gráfica (GUI) oficial para o MySQL. Ela possibilitará a criação de um banco de dados, o carregamento dos dados que serão trabalhados e a realização de consultas nos dados estruturados.

É possível obter a ferramenta MySQL Workbench através do *site* Download MySQL Workbench. Após baixá-lo, será possível abrir o arquivo executável e realizar o passo a passo de instalação que você encontrará no Vídeo 3¹⁰.

Instalados os programas, para ativar o servidor, basta abrir o XAMPP e ativar o servidor MySQL clicando no botão *Start*, como mostra a Figura 4.5.

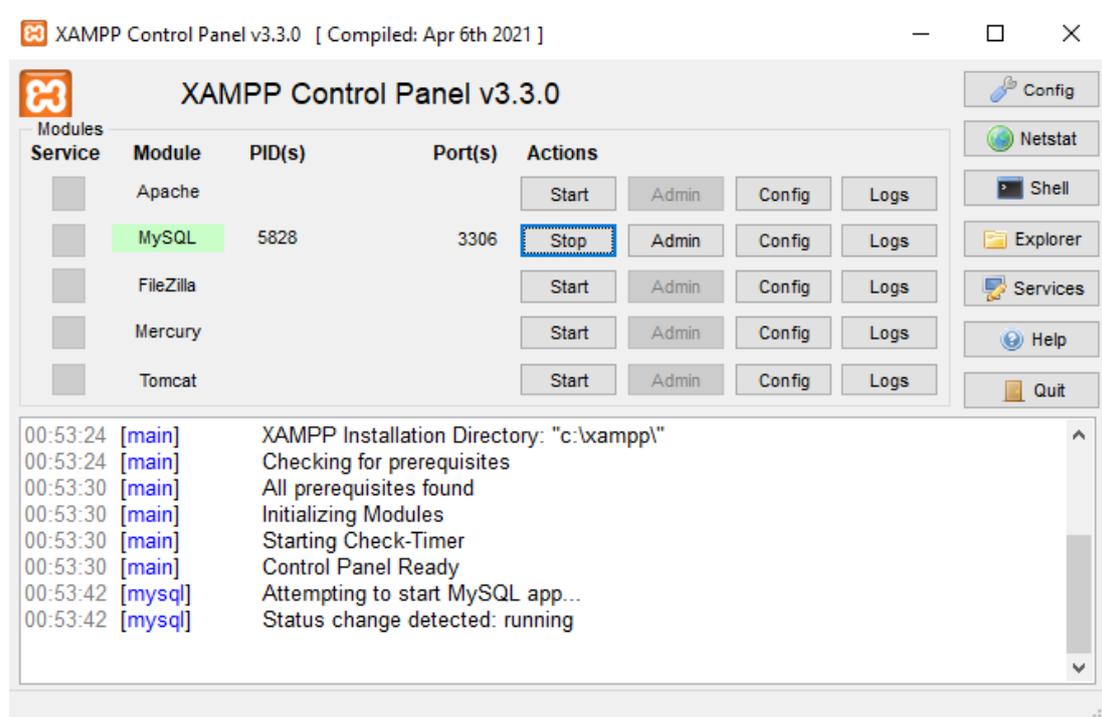


Figura 4.5. Para ativar o banco de dados MySQL, basta clicar em “*Start*”.

Depois de ativar o servidor, será o momento de criar o banco de dados para que as consultas possam ser realizadas. Todo o processo de criação do banco de dados e *upload* da tabela *mulher_2020* poderá ser acompanhado no Vídeo 4¹¹. Após isso, será possível visualizar a tabela, com as suas colunas no MySQL Workbench, tal como na Figura 4.6.

⁸ https://www.apachefriends.org/pt_br/index.html. Acesso em outubro de 2021.

⁹ O Vídeo 2 pode ser acessado em: https://youtu.be/id_DtScfJUg.

¹⁰ O Vídeo 3 pode ser acessado em <https://youtu.be/OjS9mwNskc0>.

¹¹ O Vídeo 4 pode ser acessado em <https://youtu.be/70H4XcFVQI0>.

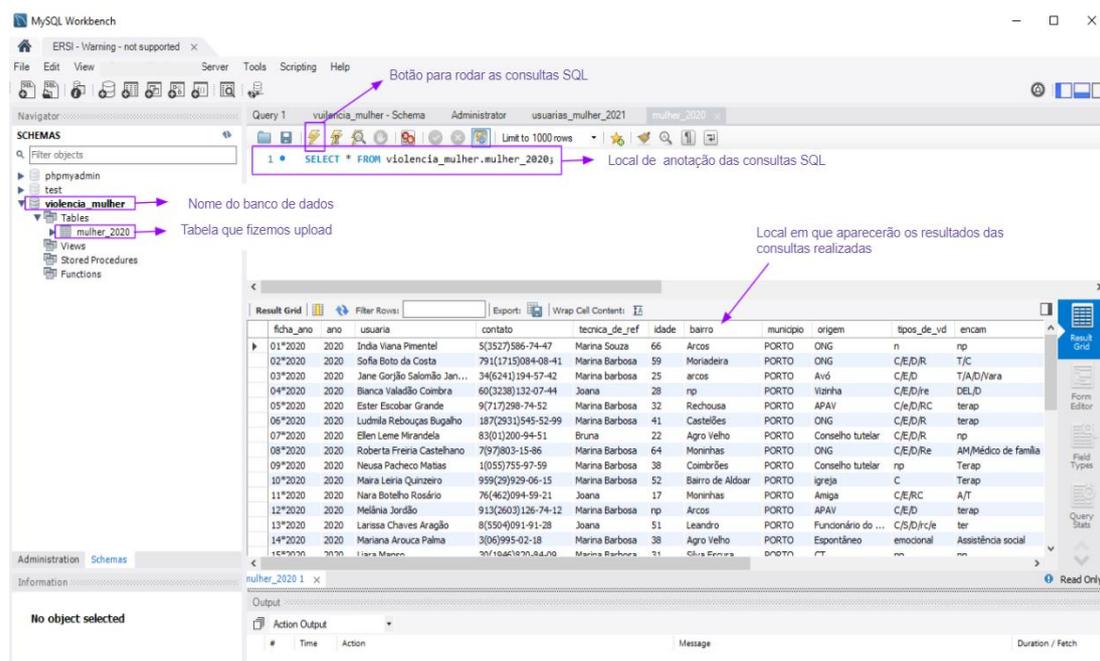


Figura 4.6. Visão geral da disposição dos elementos do MySQL Workbench que serão mais utilizados durante a realização das consultas SQL propostas neste capítulo.

Após a montagem do banco, será o momento de realizar a anonimização dos dados. Na seção 4.4 serão apresentadas algumas técnicas de anonimização de dados, além de atividades práticas para esta finalidade.

4.4. Anonimização dos Dados

A Lei 13.709 de 2018, conhecida como Lei Geral de Proteção de Dados (LGPD), prevê o tratamento dos dados pessoais de pessoa física ou jurídica de direito público ou privado para a preservação dos direitos fundamentais de liberdade, privacidade e o livre desenvolvimento da personalidade. Entre as normas gerais contidas na LGPD, a anonimização de dados que permitam identificar direta ou indiretamente um indivíduo, tem como objetivo a preservação do direito da pessoa.

Segundo a LGPD, a anonimização trata da “utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo”. Durante a análise de uma base de dados é possível haver dados pessoais que identifiquem um indivíduo, sem que essa vinculação seja desejada. Para fins de preservação, esses dados devem ser anonimizados [Brasil 2018] e, para isso, devem ser aplicadas técnicas específicas para a desvinculação dos dados sem que seja prejudicada a possibilidade de utilização da base de dados.

4.4.1. Técnicas de Anonimização

Anônimo é aquele que não apresenta nome ou assinatura, indivíduo desconhecido [Michaelis 2021]. A definição do conceito de dados pessoais pode seguir uma orientação expansionista (a partir da delimitação de “pessoa identificável”) ou reducionista (“pessoa identificada”), respectivamente alargando ou restringindo o escopo de aplicação da LGPD [Bioni 2019]. A princípio, o processo de anonimização pode estar atrelado aos dados pessoais numa orientação reducionista, contudo, é necessária uma análise contextual para verificar se na combinação de outros dados, como num quebra-cabeça, a identificação da pessoa seria possível, segundo a orientação expansionista.

Neste sentido, há técnicas que buscam eliminar elementos identificadores de uma base de dados. Bioni (2019) esclarece que o processo de anonimizar um dado é composto por técnicas que buscam eliminar elementos identificadores ou identificáveis da base em que estão organizados. O autor aponta quatro técnicas para nortear o processo: (i) a supressão; (ii) generalização; (iii) randomização e (iv) pseudoanonimização. A escolha da técnica ou da combinação delas precisa considerar os dados que devem ter seus vínculos quebrados com seus respectivos titulares.

4.4.1.1. Supressão

A técnica de supressão utiliza como recurso para a anonimização, a exclusão de campos da tabela de dados ou a substituição de parte de caracteres do campo de identificação. Em alguns casos, a substituição de caracteres pode não ser suficiente para anonimizar a base. É o caso do CPF, que é um campo que trata da identificação exclusiva de um indivíduo. Nesta situação, vale a pena considerar a exclusão do campo.

4.4.1.2. Generalização

A generalização propõe a substituição de um dado por outro que traz significado mais geral. Em algumas situações pode-se considerar a supressão como uma forma de generalização. Um exemplo é a supressão do CEP. Os últimos dígitos dos dados de CEP de uma base podem ser excluídos da mesma e ainda assim isso não ser suficiente para gerar a sua anonimização. Isto porque o CEP pode ser combinado com outros dados que fazem parte da base. Assim, a generalização propõe a substituição de um dado por outro que traga um significado mais geral, como por exemplo, a substituição do CEP pelo bairro ou a substituição da data de nascimento pela faixa etária.

4.4.1.3. Randomização

A randomização é uma técnica que busca mascarar uma informação, misturando os valores da tabela, sem afetar as análises, apenas visando não identificar seus titulares. Considerando uma tabela com a abordagem relacional, os dados de uma coluna (um campo) são alternados entre suas linhas em suas ordens, sem alterar seus valores. Por exemplo, considere a Tabela 4.1 sendo uma tabela original de dados.

Tabela 4.1. Tabela Original de Dados

Data de Nascimento	Código postal	Cor	Violência
27/05/2000	4618-4775	Preta	Psicológica
30/09/2001	8798-8956	Parda	Física
28/03/1960	4618-1228	Parda	Psicológica

Na randomização, os dados são alternados entre as linhas de registros. A Tabela 4.2 ilustra um exemplo da tabela após a randomização dos dados.

Tabela 4.2. Tabela com Dados Randomizados

Data de Nascimento	Código postal	Cor	Violência
27/05/2000	8798-8956	Parda	Psicológica
30/09/2001	4618-1228	Parda	Psicológica
28/03/1960	4618-4775	Preta	Física

É importante ter cuidado na aplicação desta técnica, pois ela pode causar impactos na análise dos dados. As respostas para perguntas que dependem da combinação entre dados de duas ou mais colunas podem ser comprometidas.

4.4.1.4. Pseudoanonimização

A pseudoanonimização, segundo a LGPD, diz respeito ao “tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro”. Nessa técnica, a base de dados é dividida em duas partes, a primeira com os dados genéricos e a segunda com os dados sensíveis.

É importante notar que não há uma técnica melhor ou combinação perfeita para direcionar o processo de anonimização, sendo essencial analisar cada contexto em que o tratamento será implementado, para que os indivíduos titulares dos dados anonimizados não sejam reidentificados.

4.4.2. Anonimizando a Base através da Pseudoanonimização

A fase de anonimização da base de dados, dependerá da forma de anonimização escolhida. Para a realização da anonimização dos dados do *dataset* CEAM, foi feita a divisão do conjunto de dados em duas tabelas (*cadastro_ceam_usuarias* e *cadastro_ceam_atual*), em um banco de dados MySQL. Para isto foi criada uma coluna com um identificador único, a fim de relacionar as tabelas.

Na tabela *cadastro_ceam_usuarias* foram reunidas 5 colunas, sendo 3 delas relativas aos dados que representam as informações julgadas como as mais sensíveis da base: nome, telefone e idade da usuária. Na tabela *cadastro_ceam_atual* foram reunidas as demais colunas constantes do *dataset* original, de modo que qualquer tarefa de mineração de dados pode ser realizada sem a necessidade de compartilhamento dos dados pessoais das usuárias. As estruturas finais das tabelas *cadastro_ceam_usuarias* e *cadastro_ceam_atual*, na base de dados já consolidada, podem ser vistas na Tabela 4.3 e na Tabela 4.4, que seguem.

Tabela 4.3. Estrutura de dados da tabela *cadastro_ceam_usuarias* com os dados pessoais das usuárias.

Name	Type	Length	Decimals	Allow Null
id	int	11	0	
ficha_ano	varchar	255	0	x
ano	varchar	255	0	x
usuaria	varchar	255	0	x
contato	varchar	255	0	x
idade	varchar	255	0	x

Tabela 4.4. Estrutura de dados da tabela *cadastro_ceam_atual* com os dados efetivamente utilizados na análise.

Name	Type	Length	Decimals	Allow Null
id	int	11	0	
ficha_ano	varchar	255	0	x
ano	varchar	255	0	x
tecnica_de_ref	varchar	255	0	x
idade	varchar	255	0	x
bairro	varchar	255	0	x
município	varchar	255	0	x
distrito	varchar	255	0	x
origem	varchar	255	0	x
tipos_de_vd	varchar	255	0	x
encam	varchar	255	0	x
dia	varchar	255	0	x
religiao_SN	varchar	255	0	x

grau_escolaridade	varchar	255	0	x
cor	varchar	255	0	x
com_renda	varchar	255	0	x
filhos	varchar	255	0	x

Para a realização da atividade proposta neste capítulo, escolheu-se criar uma tabela com as colunas que possuíam os dados pessoais das vítimas e outra tabela com os demais dados. A forma escolhida para realizar o relacionamento entre as tabelas foi através da criação de uma coluna com um identificador único (um número único para cada vítima), representada pela coluna <id>.

De forma prática, as operações que foram realizadas no banco seguiram o passo a passo descrito adiante.

4.4.2.1. Criação da coluna id na tabela *usuarias_mulher_2020*

A criação da coluna <id> é o primeiro passo importante no processo de anonimização da base, pois através da coluna <id> se pode criar um relacionamento entre a tabela com os dados pessoais das vítimas e a outra tabela, com os demais dados. Para criar a coluna <id>, será realizada a seguinte instrução SQL:

```
ALTER TABLE mulher_2020 ADD id INT NOT NULL PRIMARY KEY auto_increment;
```

Esta instrução criará a coluna <id> na tabela. Quando terminar a execução, será possível localizar a coluna <id>, que será criada após a última coluna. É possível clicar em cima do cabeçalho da coluna e arrastá-la para a esquerda, posicionando-a como a primeira coluna da tabela. É uma convenção utilizar a coluna <id> como a primeira coluna em tabelas desta finalidade, conforme mostra a Figura 4.7.

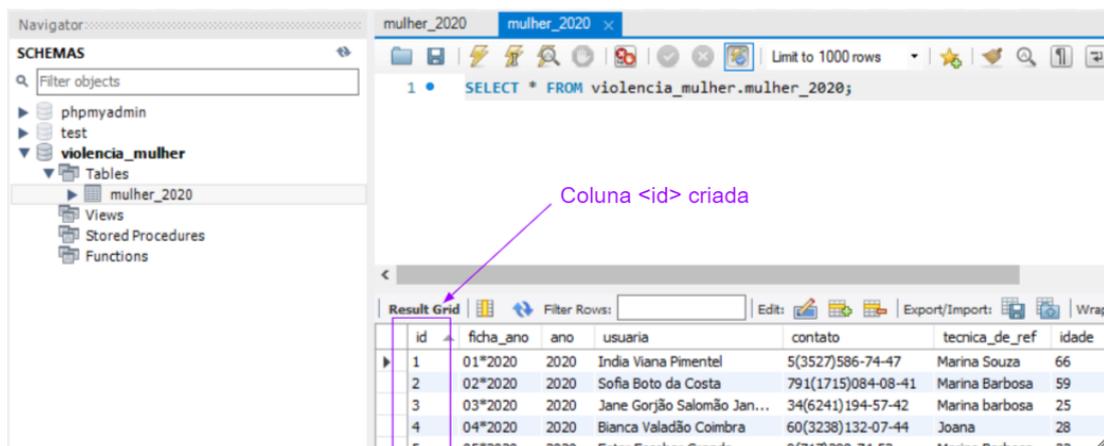


Figura 4.7. Resultado da criação da coluna <id>.

Agora que a coluna <id> está presente na tabela, é possível replicá-la a fim de criar uma tabela somente com os dados pessoais das vítimas e outra com os demais dados.

4.4.2.2. Criação da tabela *usuarias_mulher_2020* a partir da duplicação da tabela *mulher_2020* original do dataset

Para criar tabela *usuarias_mulher_2020* foi realizada a duplicação da tabela *mulher_2020* com a instrução SQL:

```
CREATE TABLE usuarias_mulher_2020 SELECT * FROM mulher_2020;
```

Esta instrução seleciona todas as colunas da tabela *mulher_2020* e cria uma tabela idêntica no banco *violencia_mulher*, nomeando-a como *usuarias_mulher_2020*. A nova tabela criada no banco aparecerá logo abaixo da tabela *mulher_2020*, conforme a Figura 4.8.

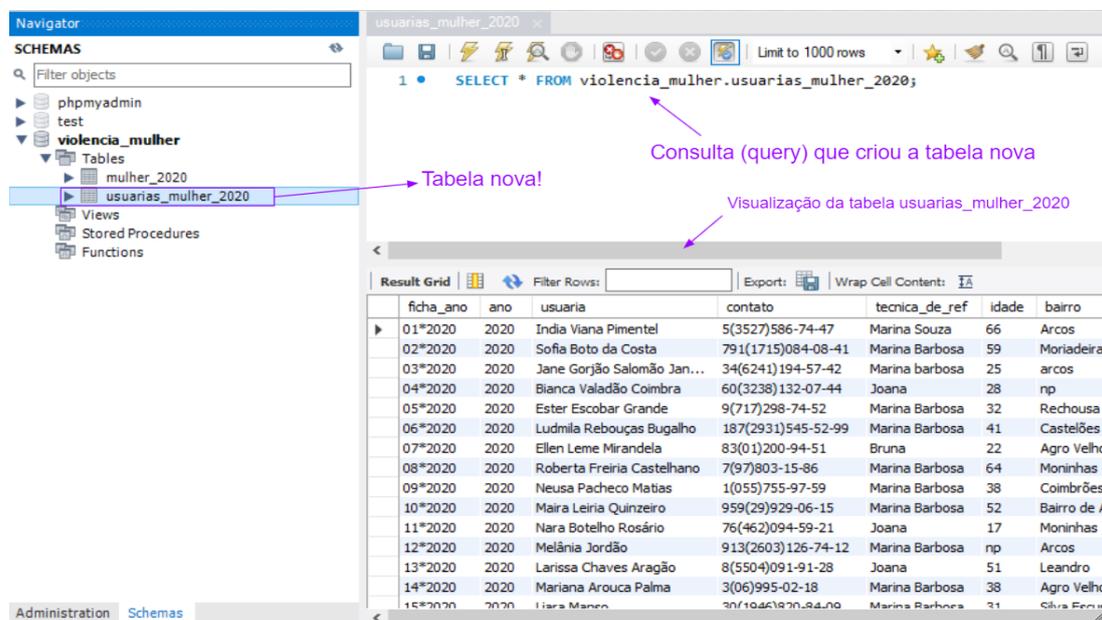


Figura 4.8. Resultado da duplicação da tabela *mulher_2020*.

Seguem também algumas recomendações em casos de erro:

1. Caso a tabela *usuarias_mulher_2020* não apareça, talvez os dados não tenham sido atualizados automaticamente. Para isso, clique com o botão direito do mouse no banco <violencia_mulher> e selecione a opção <Refresh All>. Faça o mesmo em <Tables>, caso ainda assim a tabela nova não apareça.
2. Para visualizar a tabela *usuarias_mulher_2020*, basta clicar no ícone em formato de tabela, que se encontra ao lado do nome da tabela, tal qual está indicado na Figura 4.9.

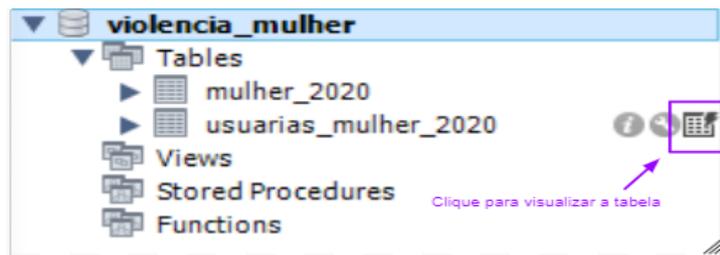


Figura 4.9. Ícone para visualização da tabela.

4.4.2.3. Organização da tabela *mulher_2020* através da supressão das colunas da tabela *usuarias_mulher_2020*, exceto as colunas <id>, <ficha_ano> e <ano>

A supressão das colunas com os dados pessoais da tabela *mulher_2020* foi realizada considerando que os dados pessoais eram os das colunas <usuaria>, <contato>, <idade> e que as colunas <id>, <ficha_ano> e <ano> deveriam permanecer tanto na tabela *mulher_2020* quanto na tabela *usuarias_mulher_2020*. As colunas suprimidas a fim de anonimizar a tabela *mulher_2020* foram <usuaria>, <contato> e <idade>. Para a supressão de colunas em uma tabela pode ser usada a estrutura:

```
ALTER TABLE nome da tabela DROP COLUMN coluna a ser suprimida;
```

Esta instrução SQL altera a tabela *mulher_2020*, retirando as colunas indicadas. Desta forma, foi realizada a instrução:

```
ALTER TABLE mulher_2020
  DROP COLUMN usuaria,
  DROP COLUMN contato,
  DROP COLUMN idade;
```

Após a supressão das colunas com os dados pessoais da tabela *mulher_2020*, as colunas que não foram consideradas dados pessoais na tabela *usuarias_mulher_2020* também foram suprimidas.

4.4.2.4. Organização da tabela *usuarias_mulher_2020* através da supressão das colunas com os dados que não foram considerados dados pessoais

As colunas suprimidas foram: <tecnica_de_ref>, <bairro>, <municipio>, <origem>, <tipos_de_vd>, <encam>, <dia>, <religiao>, <grau_escolaridade>, <cor>, <com_renda>, <filhos> e <rel_agressor>. A supressão das colunas foi feita através da consulta:

```
ALTER TABLE usuarias_mulher_2020
DROP COLUMN tecnica_de_ref,
DROP COLUMN bairro,
DROP COLUMN municipio,
DROP COLUMN origem,
DROP COLUMN tipos_de_vd,
DROP COLUMN encam,
DROP COLUMN dia,
DROP COLUMN religiao,
DROP COLUMN grau_escolaridade,
DROP COLUMN cor,
DROP COLUMN com_renda,
DROP COLUMN filhos,
DROP COLUMN rel_agressor;
```

A partir da supressão das colunas desejadas, obteve-se então a consolidação da tabela *usuarias_mulher_2020*, que figura na base de dados com o objetivo de identificar as usuárias somente caso necessário, viabilizando o sigilo dos dados pessoais das vítimas durante as análises de dados realizadas com a tabela *mulher_2020*, conforme mostra a Figura 4.10.



id	ficha_ano	ano	usuaria	contato	idade
1	01*2020	2020	India Viana Pimentel	5(3527)586-74-47	66
2	02*2020	2020	Sofia Boto da Costa	791(1715)084-08-41	59
3	03*2020	2020	Jane Gorjão Salomão Jan...	34(6241)194-57-42	25
4	04*2020	2020	Bianca Valadão Coimbra	60(3238)132-07-44	28
5	05*2020	2020	Ester Escobar Grande	9(717)298-74-52	32
6	06*2020	2020	Ludmila Rebouças Bugalho	187(2931)545-52-99	41
7	07*2020	2020	Ellen Leme Mirandela	83(01)200-94-51	22
8	08*2020	2020	Roberta Freiria Castelhana	7(97)803-15-86	64
9	09*2020	2020	Neusa Pacheco Matias	1(055)755-97-59	38

Figura 4.10. Tabela *usuarias_mulher_2020* consolidada, realizando a guarda das informações que devem ser mantidas em sigilo.

Finalizada a etapa de anonimização dos dados, através da separação dos dados pessoais das vítimas dos demais dados, inicia-se a etapa de limpeza dos dados.

4.4.3. Limpeza dos Dados

O processo de limpeza de dados é uma parte integrante da análise de dados como um todo e representa o processo de fixar ou remover dados formatados incorretamente, corrompidos, duplicados ou incompletos.

Durante a análise exploratória dos dados contidos na tabela *mulher_2020*, constata-se o uso de múltiplos valores para a representação de uma mesma informação dentro de uma determinada coluna da tabela, como por exemplo: a existência dos valores “ca.”, “católica”, “catolica”, “catól”, “CAT”, “catolica” e outros para representar a religião “católica” na coluna religião da tabela *mulher_2020*. Esse problema deve-se ao

fato de que o *dataset* fictício que utilizamos para este minicurso foi criado com base no *dataset* real da CEAM, em que a anotação das informações das mulheres vítimas de violência foi realizada de forma manual, por funcionários de uma instituição, ocasionando problemas de digitação e ausência de padronização dos valores.

Na experiência realizada com o *dataset* do CEAM, para resolver esse problema da padronização dos valores e garantir que a base pudesse retornar consultas SQL de modo eficiente, foi necessário realizar alguns procedimentos para a normalização dos valores das colunas da tabela. Pode-se visualizar na Figura 4.11 os diferentes valores utilizados para representar a informação da religião da usuária.



Figura 4.11. Valores distintos encontrados inicialmente na coluna “religião” do *dataset* CEAM, antes da limpeza de dados (à esquerda) e depois (à direita).

Na construção da base de dados da CEAM, o procedimento de normalização consistiu primeiro em identificar quais foram os valores distintos, usados na representação de um mesmo dado, estavam presentes em cada uma das colunas da tabela, para depois então passar à padronização desses dados. Esse procedimento foi repetido para todas as colunas da base, resultando em diminuição substancial dos valores distintos em cada uma das colunas, como se pode observar na Tabela 4.5 e na Figura 4.12.

Tabela 4.5. Redução substancial dos valores distintos para a representação dos dados existentes nas colunas da tabela *cadastro_ceam_atual* após a limpeza de dados.

	Quantidade de valores distintos encontrados nas colunas						
	origem	bairro	tipos_de_vd	encam	religiao	grau_escolaridade	cor
Antes da limpeza	140	71	150	189	43	21	21
Depois da limpeza	123	43	29	86	13	10	10
Redução	12%	39%	81%	54%	70%	52%	52%

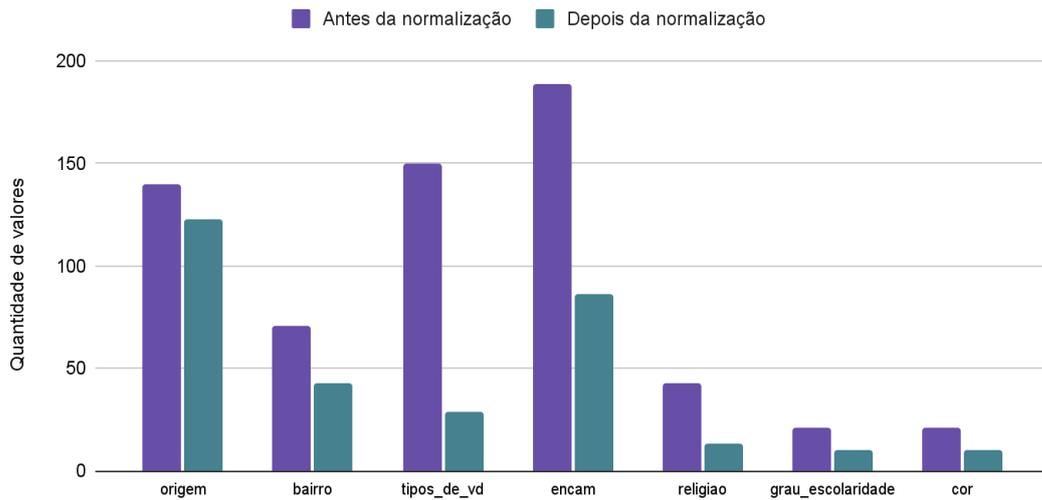


Figura 4.12. Relação entre os valores das colunas antes e depois dos procedimentos adotados para a normalização.

A transformação dos dados no contexto apresentado esteve relacionada à uma característica marcante do *dataset* real analisado, que foi a falta de padronização dos dados, em decorrência da anotação manual dos dados pela equipe técnica responsável pelos registros. A fim de tornar a tabela *mulher_2020* em um objeto próprio para a realização de consultas, procedeu-se então à padronização dos dados, através dos procedimentos descritos nas duas próximas subseções.

4.4.3.1. Identificação dos valores distintos em uma determinada coluna da tabela *mulher_2020*

A identificação dos valores distintos em uma determinada coluna, foi realizada com a instrução:

```
SELECT <coluna>, COUNT(1) AS qtd FROM mulher_2020 GROUP BY
<coluna> ORDER BY qtd DESC, <coluna>;
```

Esta consulta conta o número de ocorrências de cada valor distinto presente em uma determinada coluna da tabela. A consulta retorna esses valores distintos agrupados por ocorrência em uma coluna e em outra coluna, quantas vezes cada valor distinto apareceu. Por exemplo, ao realizar a consulta feita anteriormente na tabela *mulher_2020*:

```
SELECT religiao, COUNT(1) AS qtd FROM mulher_2020 GROUP BY
religiao ORDER BY qtd DESC, religiao;
```

O uso dessa consulta retorna uma listagem com duas colunas. Uma delas com todos os diferentes dados relativos às religiões das vítimas constantes na coluna <religiao>, ao lado da contagem da incidência de cada um dos tipos, tal como na Figura 4.13.

religiao	qtd
atéia	3
CAT	3
budista	2
afro	1
bud/xinto	1
ca.	1
calvinista	1
cat.	1
catol	1
catól.	1
CATOLIC	1
lut.	1
m afro	1
M. Afro	1
matriz afro	1
mussulmana	1
não	1
testemun...	1
xintoista	1

← Valores distintos presentes na coluna "religiao"

Figura 4.13. Valores distintos presentes na coluna "religião" da tabela *mulher_2020*.

4.4.3.2. Criação de um padrão correspondente ao dado que se desejava representar e padronização dos dados referentes

Esta etapa da limpeza de dados foi a que permitiu efetivamente que se pudesse trabalhar com os dados em uma fase de análise. Uma vez encontrados os valores distintos possíveis para representar um mesmo dado na tabela *mulher_2020*, criou-se um padrão para representar cada um dos valores. Por exemplo, onde se lia “CAT”, “católica”, “ca.”, optou-se pelo padrão “catolica”. Depois, passou-se para a etapa de padronização em si. Nessa etapa, utilizou-se a instrução SQL a seguir tantas vezes quantas foram necessárias para a padronização da nomenclatura de um determinado dado:

```
UPDATE mulher_2020 SET <coluna> = dado padronizado
WHERE <coluna> = dado não padronizado;
```

Essa instrução basicamente substitui um determinado valor presente em uma coluna da tabela especificada por outro valor informado. Ao realizar as instruções UPDATE a seguir, serão padronizados os valores “CAT”, “CATOLICA”, “CATÓLICA” e “catól.” como “catolica”.

```
UPDATE  mulher_2020  SET  religiao  =  'catolica'
WHERE  religiao  =  'CAT';

UPDATE  mulher_2020  SET  religiao  =  'catolica'
WHERE  religiao  =  'CATOLICA';

UPDATE  mulher_2020  SET  religiao  =  'catolica'
WHERE  religiao  =  'CATÓLICA';

UPDATE  mulher_2020  SET  religiao  =  'catolica'
WHERE  religiao  =  'catól.';
```

Essa ação viabiliza a contagem desses valores pelas consultas SQL, uma vez que os dados estarão padronizados.

Atenção:

- Sempre que se utilizar palavras como dados da tabela durante a consulta SQL, elas deverão vir entre aspas simples (' ').
- O MySQL Workbench possui uma configuração inicial que impede as instruções que usam os comandos UPDATE e DETETE. Para configurar o sistema para aceitar essas instruções você deverá ir em <edit><preferences> e desmarcar a caixa de seleção de “safe updates”, como mostra a Figura 4.14.

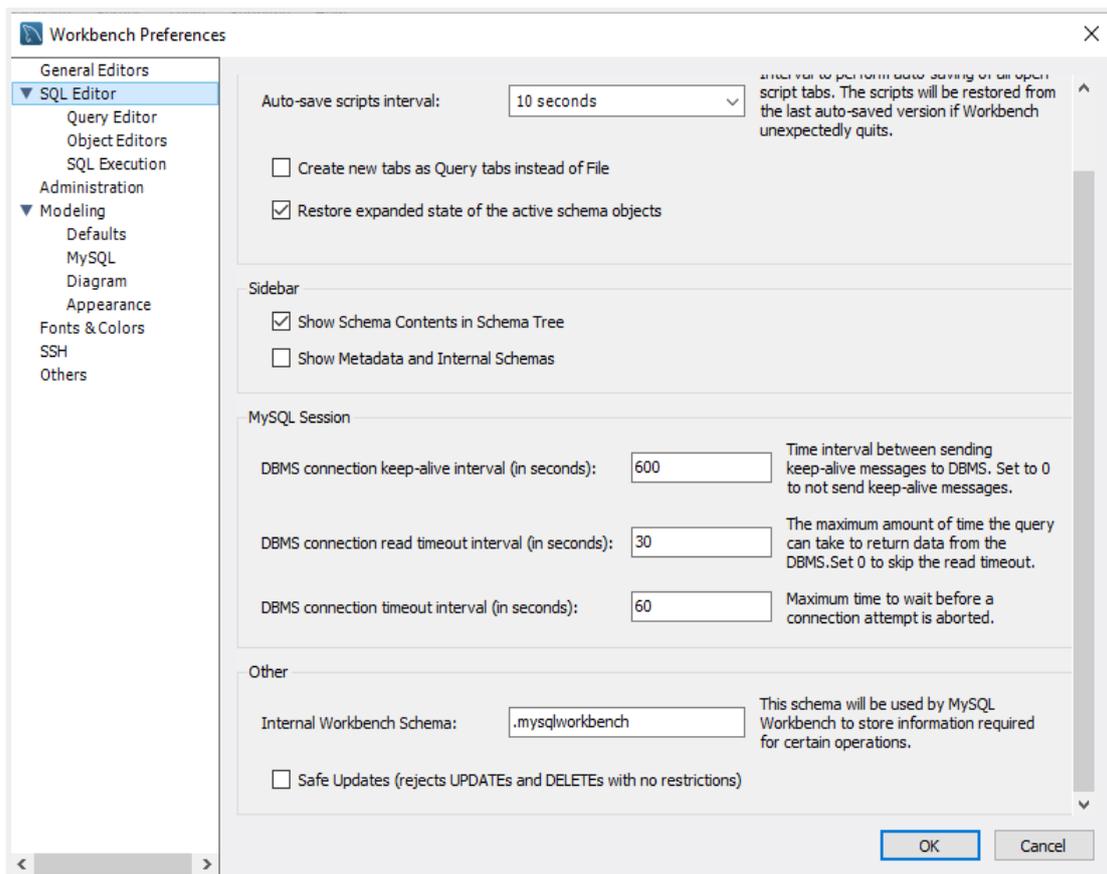


Figura 4.14. A caixa de seleção “safe updates” deve ser desmarcada para a realização do comando update.

Após o procedimento de normalização, espera-se obter uma tabela mais apropriada para a realização de análises e dados e criação de visualizações.

4.4.4. Transformando os Dados

Finalizando as etapas referentes à consolidação da base de dados para a fase de análise, é muito importante que as colunas exibam dados monovalorados. Um problema importante do *dataset mulher_2020* está na coluna <tipos_de_vd>, que relaciona os tipos de violência praticados pelos agressores, apresentando dados multivalorados, ao invés de monovalorados.

A coluna <tipos_de_vd> pode apresentar até 5 valores dentro de uma mesma célula, como por exemplo, “CORP/EMO/DIG/REC/SEX”, referentes às violências “corporal”, “emocional”, “dignidade”, “recursos/patrimônio” e “sexual”. Essa disposição da informação pode dificultar muito o processamento dos dados quando carregados em um *software* de visualização de dados.

O processo de transformação dos dados representa a tarefa de modificar o formato de dados para um mais interessante ou apropriado para a mineração de dados, de maneira a agilizar o processamento de dados em bases mais extensas.

Para tornar as colunas multivaloradas em monovaloradas, foi adotada a estratégia de transformação de uma coluna multivalorada em várias colunas monovaloradas, contendo como dados 0 ou 1 (type int). No *dataset mulher_2020*, as colunas que podem receber esse tipo de transformação são a coluna <tipos_de_vd> e a coluna <encam>, que guardam os dados dos tipos de violência doméstica praticados e dos locais para onde as vítimas foram encaminhadas no momento da denúncia.

As colunas novas, com dados monovalorados, serão proporcionais aos valores existentes na coluna com dados multivalorados. Por exemplo, se a coluna <tipos_de_vd> pode apresentar os dados CORP/EMO/DIG/REC/SEX, é possível transformá-la em 5 novas colunas, as colunas <CORP>, <EMO>, <DIG>, <REC> e <SEX>, assumindo valor “0” caso o tipo de violência não esteja presente no cadastro e “1” caso o tipo de violência esteja presente no cadastro.

Para realizar a transformação, deve-se seguir os passos:

1. Identificar as colunas que possuem os dados multivalorados;
2. Identificar quais são os dados possíveis dentro de uma coluna de dados multivalorados;
3. Criar tantas novas colunas na tabela quantas forem as possibilidades de valores identificados no passo anterior;
4. Uniformização dos nomes representativos dos dados na coluna <tipos_de_vd>;
5. Representar a ocorrência (1) e a não ocorrência (0) de cada um dos dados nas novas colunas criadas.

4.4.4.1. Identificação das colunas que possuíam dados multivalorados

A identificação das colunas que possuem dados multivalorados deve se dar através da leitura livre dos dados das colunas da tabela *mulher_2020* e da percepção de dados escritos lado a lado ou mesmo separados por vírgulas, ponto e vírgula ou barras.

4.4.4.2. Identificação e anotação de todos os dados possíveis dentro de uma mesma coluna

Na coluna <tipos_de_vd>, pode se observar a olho nu, que existem 5 possibilidades de violência sofrida “CORP”, “EMO”, “DIG”, “REC” e “SEX”.

4.4.4.3. Criação das novas colunas “corp”, “emo”, “dig”, “rec”, “sex” na tabela, representando as possibilidades de valores identificados na coluna multivalorada

Para criar as novas colunas aplica-se a instrução:

```
ALTER TABLE mulher_2020 ADD coluna x INT AFTER coluna y;
```

Essa instrução SQL criará uma coluna de nome <coluna x> logo após a <coluna y> informada. Para criar as colunas desejadas na tabela, basta executar as instruções que seguem. A estrutura resultante é mostrada na Figura 4.15.

```
ALTER TABLE mulher_2020
ADD corp int AFTER rel_agressor,
ADD emo int AFTER corp,
ADD dig int AFTER emo,
ADD rec int AFTER dig,
ADD sex int AFTER rec;
```

cam	dia	religiao	grau_escolaridade	cor	com_renda	filhos	rel_agressor	corp	emo	dig	rec	sex
	1/2/2020	np	ensino médio	negra	não	1	Ex marido	NULL	NULL	NULL	NULL	NULL
	1/2/2020	CAT	ensino fundamental	negra	sim	3	outros	NULL	NULL	NULL	NULL	NULL
/D/Vara	1/3/2020	luter	alfabetização	branca	sim	8	marido	NULL	NULL	NULL	NULL	NULL
/D	1/4/2020	catolica	ensino médio	amarela	não	1	irmã	NULL	NULL	NULL	NULL	NULL
ap	1/4/2020	calvinista	ensino médio	negra	sim	2	marido	NULL	NULL	NULL	NULL	NULL
ap	1/5/2020	np	np	branca	sim	3	mãe	NULL	NULL	NULL	NULL	NULL
	1/8/2020	bud/xinto	ensino médio	amarela	sim	3	marido	NULL	NULL	NULL	NULL	NULL
/Médico de família	1/8/2020	budista	alfabetização	amarela	sim	0	desconhecido	NULL	NULL	NULL	NULL	NULL
ap	1/10/2020	atéia	ensino fundamental	mestiça	sim	2	nora	NULL	NULL	NULL	NULL	NULL
ap	1/12/2020	CATOLICA	alfabetização	amarela	sim	1	np	NULL	NULL	NULL	NULL	NULL
	1/15/2020	np	alfabetização	negra	np	7	marido	NULL	NULL	NULL	NULL	NULL
ap	1/16/2020	CA	alfabetização	mestiça	sim	2	marido	NULL	NULL	NULL	NULL	NULL
	1/22/2020	M. Afro	ensino médio	negra	sim	4	companheiro	NULL	NULL	NULL	NULL	NULL
istência social	1/23/2020	ca.	ensino médio	branca	não	2	companheiro	NULL	NULL	NULL	NULL	NULL
	1/23/2020	CATÓLICA	np	branca	não	3	tio	NULL	NULL	NULL	NULL	NULL

Figura 4.15. Colunas “corp”, “emo”, “dig”, “rec” e “sex”, criadas após a coluna *rel_agressor*.

4.4.4.4. Colocando todos os dados da coluna tipos_de_vd em letra maiúscula

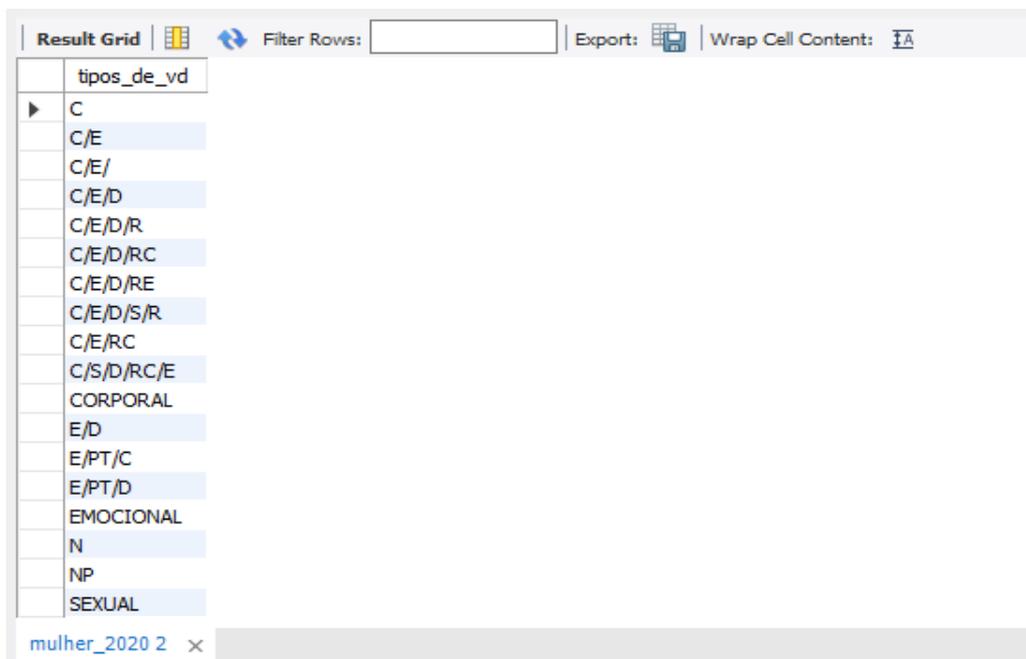
Para tratar na padronização da representação dos dados na coluna, a fim de facilitar as próximas transformações que serão feitas na tabela, é indicado colocar todos os dados em caixa alta ou em caixa baixa. Para colocá-los em caixa alta, realiza-se a seguinte instrução:

```
UPDATE mulher_2020 SET tipos_de_vd = UPPER(tipos_de_vd);
```

4.4.4.5. Identificando os diferentes possíveis valores para a coluna tipos_de_vd e instanciação das colunas geradas

O próximo passo é descobrir quais são os valores distintos possíveis na coluna <tipos_de_vd>. Isto porque, a partir dessa descoberta, será possível compreender quais novas instruções serão necessárias para levar a informação da coluna multivalorada <tipos_de_vd> para as novas colunas criadas. Para saber quais são os valores distintos da coluna <tipos_de_vd>, realiza-se a seguinte consulta:

```
SELECT DISTINCT tipos_de_vd FROM mulher_2020 order by tipos_de_vd;
```



tipos_de_vd
C
C/E
C/E/
C/E/D
C/E/D/R
C/E/D/RC
C/E/D/RE
C/E/D/S/R
C/E/RC
C/S/D/RC/E
CORPORAL
E/D
E/PT/C
E/PT/D
EMOCIONAL
N
NP
SEXUAL

Figura 4.16. Distintas possibilidades de valores na coluna <tipos_de_vd>.

A partir da identificação dos valores possíveis na coluna <tipos_de_vd>, a ação seguinte se concentra em preencher as colunas geradas com 1, iniciando pela coluna <corp>. Observando a Figura 4.16, percebe-se que todas as células que possuem a informação de que houve violência corporal, possuem a letra “C” como indicação. Entretanto, como a letra “C” está vindo seguida de diversas informações dentro das células, será necessário distinguir que só queremos colocar 1 na coluna <corp> para as células cuja informação se inicia com a letra “C”, podendo ser seguida de qualquer outro

dado. Para realizar essa representação, utiliza-se o símbolo de porcentagem (%) logo após a letra “C”. Assim: ‘C%’, não esquecendo das aspas simples para indicar que se trata de uma *string*.

```
UPDATE mulher_2020 SET corp = 1 WHERE tipos_de_vd LIKE 'C%';
```

Semelhante modo deve ser feito para preencher a coluna <corp> com 1 quando o dado da coluna <tipo_de_vd> for igual a *string* ‘CORPORAL’.

```
UPDATE mulher_2020 SET corp = 1 WHERE tipos_de_vd LIKE 'CORPORAL';
```

Procedimentos semelhantes serão aplicados para a construção das outras colunas. Entretanto, é importante observar que a análise para a definição da melhor instrução deverá ser feita caso a caso.

Para preencher a coluna <emo> com 1 quando o dado na coluna <tipos_de_vd> começar com a letra “E”, contiver a letra “E” entre duas barras (/), contiver a string “EMOCIONAL” ou terminar com uma barra (/) seguida da letra “E” usa-se as instruções:

```
UPDATE mulher_2020 SET emo = 1 WHERE tipos_de_vd LIKE 'E%';
UPDATE mulher_2020 SET emo = 1 WHERE tipos_de_vd LIKE '%/E/%';
UPDATE mulher_2020 SET emo = 1 WHERE tipos_de_vd LIKE 'EMOCIONAL';
UPDATE mulher_2020 SET emo = 1 WHERE tipos_de_vd LIKE '%/E';
```

Segue a mesma lógica para preencher a coluna <dig> com 1, quando o dado na coluna <tipos_de_vd> terminar com “/D” ou possuir a letra “D” entre duas barras (/D/).

```
UPDATE mulher_2020 SET dig = 1 WHERE tipos_de_vd LIKE '/D';
UPDATE mulher_2020 SET dig = 1 WHERE tipos_de_vd LIKE '/D/%';
```

As instruções que seguem efetivam o preenchimento da coluna <rec> com 1, quando o dado na coluna <tipos_de_vd> terminar com “/R”, terminar com “/RC”, terminar com “/RE”, contiver “/RC/” e quanto contiver “/PT/”.

```
UPDATE mulher_2020 SET rec = 1 WHERE tipos_de_vd LIKE '/R';
UPDATE mulher_2020 SET rec = 1 WHERE tipos_de_vd LIKE '/RC';
UPDATE mulher_2020 SET rec = 1 WHERE tipos_de_vd LIKE '/RE';
UPDATE mulher_2020 SET rec = 1 WHERE tipos_de_vd LIKE '/RC/%';
UPDATE mulher_2020 SET rec = 1 WHERE tipos_de_vd LIKE '/PT/%';
```

Para preencher a coluna <sex> com 1, quando o dado na coluna <tipos_de_vd> for igual à string “SEXUAL” use a próxima instrução:

```
UPDATE mulher_2020 SET sex = 1 WHERE tipos_de_vd LIKE 'SEXUAL';
```

Depois que terminar essa etapa de preenchimento das colunas, elas terão os valores “1” e “null” como dados. Então será o momento de trocar os valores “null” por “0”, para obter o formato desejado no planejamento da transformação de dados. As próximas instruções de UPDATE realizam a alteração desejada.

```
UPDATE mulher_2020 SET corp = 0 WHERE corp is null;
```

```
UPDATE mulher_2020 SET emo = 0 WHERE emo is null;
```

```
UPDATE mulher_2020 SET dig = 0 WHERE dig is null;
```

```
UPDATE mulher_2020 SET rec = 0 WHERE rec is null;
```

```
UPDATE mulher_2020 SET sex = 0 WHERE sex is null;
```

O resultado dessas consultas podem ser visualizados na Figura 4.17.

id_vd	encam	dia	religiao	grau_escolaridade	cor	com_renda	filhos	rel_agressor	corp	emo	dig	rec	sex
	np	1/2/2020	np	ensino médio	negra	não	1	Ex marido	0	0	0	0	0
	T/C	1/2/2020	CAT	ensino fundamental	negra	sim	3	outros	1	1	1	1	0
	T/A/D/Vara	1/3/2020	luter	alfabetização	branca	sim	8	marido	1	1	1	0	0
E	DEL/D	1/4/2020	catolica	ensino médio	amarela	não	1	irmã	1	1	1	1	0
C	terap	1/4/2020	calvinista	ensino médio	negra	sim	2	marido	1	1	1	1	0
	terap	1/5/2020	np	np	branca	sim	3	mãe	1	1	1	1	0
	np	1/8/2020	bud/xinto	ensino médio	amarela	sim	3	marido	1	1	1	1	0
E	AM/Médico de família	1/8/2020	budista	alfabetização	amarela	sim	0	desconhecido	1	1	1	1	0
	Terap	1/10/2020	atéia	ensino fundamental	mestiça	sim	2	nora	0	0	0	0	0
	Terap	1/12/2020	CATOLICA	alfabetização	amarela	sim	1	np	1	0	0	0	0
	A/T	1/15/2020	np	alfabetização	negra	np	7	marido	1	1	0	1	0
	terap	1/16/2020	CA	alfabetização	mestiça	sim	2	marido	1	1	1	0	0
C/E	ter	1/22/2020	M. Afro	ensino médio	negra	sim	4	companheiro	1	1	1	1	0
NAL	Assistência social	1/23/2020	ca.	ensino médio	branca	não	2	companheiro	0	1	0	0	0
	np	1/23/2020	CATOLICA	np	branca	não	3	tio	0	0	0	0	0

Figura 4.17. Formato das colunas “corp”, “emo”, “dig”, “rec” e “sex” depois do processo de transformação.

4.4.5. Resultados da Base Consolidada

Como resultado do processo de aquisição e transformação dos dados, obteve-se a base de dados nomeada *violencia_mulher*, com as tabelas *mulher_2020* e *usuarias_mulher_2020*, relacionadas através da coluna <id>. A base consolidada permite a etapa da integração de dados, que é o carregamento de dados, que tem como objetivo gerar valor sobre eles para a formação de relatórios e análises capazes de dar suporte à tomada de decisão. Um exemplo de etapa de integração será mostrado na seção a seguir, com a utilização da base consolidada para a criação de um *dashboard* interativo.

4.5. Sumarização de Dados com *Dashboards*

Nesta seção, serão descritas algumas técnicas e práticas relacionadas à sumarização de dados através de *dashboards*, passando pela criação de perguntas de pesquisa a serem feitas diante da base de dados consolidada até as técnicas geração de *dashboards* e as decisões de *design* que permeiam todo o processo.

4.5.1. Como Visualizar os Dados?

A partir da base gerada dos dados, a análise dos dados buscou responder questionamentos que as especialistas do CEAM precisam submeter em relatórios, a fim de simular perguntas de pesquisa que dariam suporte para gestores/tomadores de decisão no âmbito das políticas públicas para população em estado de vulnerabilidade social — Mulheres. Desta forma, foram levantadas as perguntas:

- Qual o número de registros coletados?
- Qual é a idade média das vítimas?
- Qual é o número médio de filhos das vítimas?
- Qual é o perfil da distribuição da quantidade de filhos das vítimas de agressão?
- Qual é a proporção entre as raças das usuárias?
- Quais são os graus de parentesco entre as vítimas e os agressores?
- Quais são os tipos de agressões mais praticadas? (não deixando de considerar que a base de dados informa que as vítimas informam na grande maioria das vezes sofrer mais de um tipo de violência, de forma cumulativa).
- Qual é a distribuição da quantidade de vítimas agredidas por distritos do município?
- Qual é a incidência de agressões considerando a divisão entre mulheres que possuem renda e as que não possuem renda?

Na próxima seção, serão realizadas análises a fim de encontrar respostas para as perguntas listadas. Através da elaboração de um *dashboard* para visualização dos dados, serão apresentadas as ferramentas e metodologias que podem contribuir para a realização da sumarização de dados e informação dos resultados das análises para a audiência.

4.5.2. Decisões Técnicas

A forma de sumarização dos dados escolhida para criar uma visualização com os dados da CEAM foi a criação de um *dashboard*, de maneira que as perguntas de pesquisa levantadas na seção 4.4.1 puderam ser filtradas e cruzadas. Esse *dashboard* foi concebido através da ferramenta Google Data Studio, que possibilita um ambiente colaborativo diretamente integrado com a plataforma Google Sheets, para onde os dados poderão ser posteriormente exportados.

Os dados disponibilizados no *dashboard* foram separados em duas páginas, a primeira contendo as informações principais discutidas na seção 4.4.1, e a segunda

possuindo um enfoque maior nos tipos de violência ocorridos. Para facilitar a análise através de alguns critérios, foram implementados filtros como o ano da agressão e a raça da mulher agredida.

4.5.3. Gerando o *Dashboard*

Para gerar um ambiente que proporcione a visualização dos dados, é necessário gerar um novo relatório na ferramenta Google Data Studio. Após fazer o login com uma conta Google, todos os seus relatórios podem ser acessados através da página Google Data Studio¹².

Uma das grandes vantagens da ferramenta Google Data Studio é a integração facilitada com os dados presentes na ferramenta de armazenamento da nuvem Google Drive. Selecionando a opção “*Blank Report*”, é possível importar a base de dados gerada, que deverá ser armazenada em uma planilha na ferramenta Google Sheets, conforme a Figura 4.18.

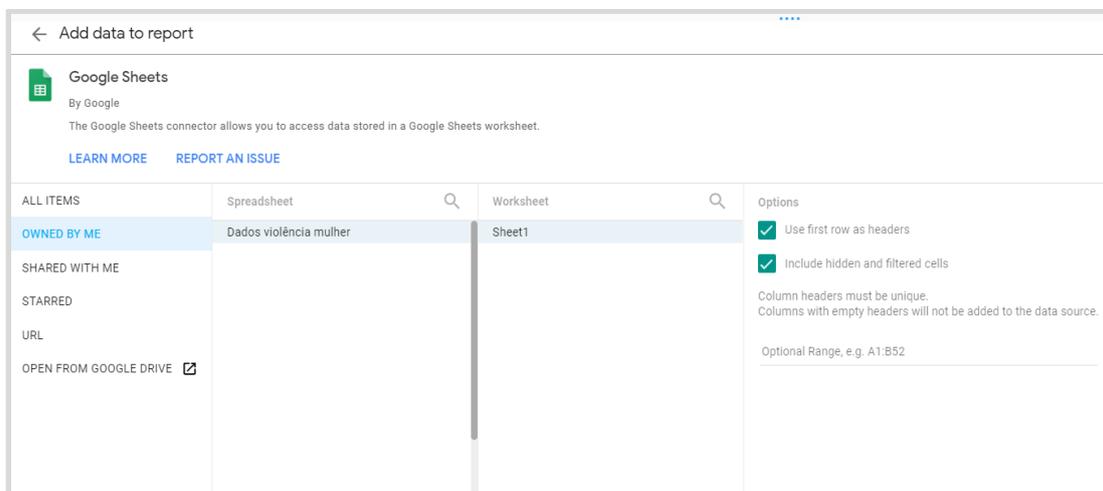


Figura 4.18. Importação da base de dados na ferramenta Google Data Studio.

Após importar a base de dados, será possível inserir a primeira visualização. Para saber a quantidade total de registros coletados, será gerada uma visualização de *scorecard* para mostrar estes dados. Essas opções de visualização podem ser vistas a partir da opção “*Add a Chart*”, localizada no menu superior da ferramenta, conforme a Figura 4.19.

¹² <https://datastudio.google.com/>. Acesso em 29 de agosto de 2021.

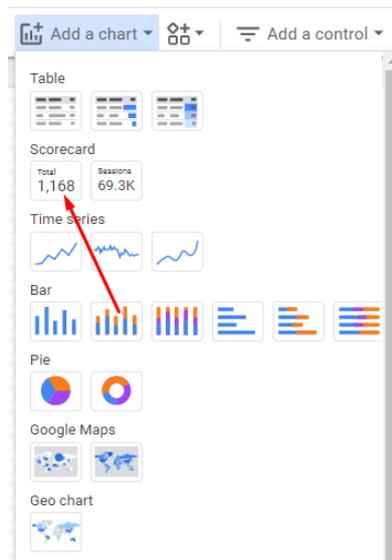


Figura 4.19. Opções de visualização.

Ao gerar uma visualização, um menu à direita será disponibilizado, mostrando alguns dos atributos principais que podem ser alterados para que sejam fornecidas as informações desejadas. Estes atributos estão presentes na Figura 4.20, seguida de uma breve descrição das suas características individuais principais.

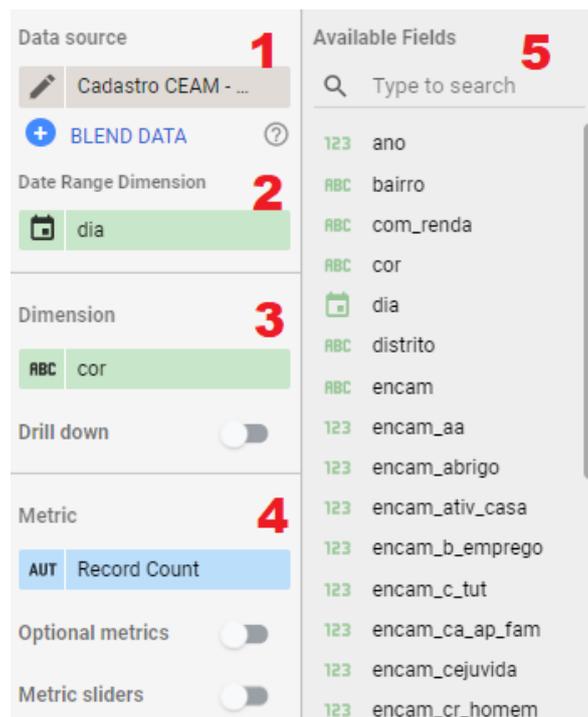


Figura 4.20. Os atributos principais em uma visualização de dados no Google Data Studio: *Data Source*, *Date Range Dimension*, *Dimension*, *Metric* e *Available Fields*.

4.5.3.1. *Data Source*

Representa a fonte de dados, ou seja, a base a partir da qual será obtida essa métrica. Caso a análise seja feita olhando em duas bases ou mais, é importante sempre verificar qual é a base a ser consultada.

4.5.3.2. *Date Range Dimension*

A dimensão de intervalo temporal informa ao Data Studio quais das dimensões disponíveis controlam o intervalo de tempo a ser trabalhado. Normalmente, este campo é preenchido automaticamente por campos que possuem a tipagem de data, por conseguirem auxiliar a determinar um intervalo para a análise.

4.5.3.3. *Dimension*

Este campo determina a dimensão na qual as métricas serão baseadas. Uma dimensão serve para descrever e categorizar os seus dados. Dessa forma, os dados serão agrupados dado algum campo único que será utilizado, garantindo a individualidade dos registros.

4.5.3.4. *Metric*

Este campo determina e mensura as métricas que serão analisadas na visualização. Uma “métrica” é o resultado de uma aplicação ou agregação de um conjunto de valores. Dentro deste campo, é possível fazer análises de seus valores através de algumas medidas matemáticas, como somatório, média ou número total de registros.

Neste campo, também é possível editar o nome do campo a ser mostrado na métrica, clicando no ícone de editar, logo acima da métrica. Isto será especialmente útil para inserir nomes mais intuitivos na visualização, conforme a Figura 4.21.

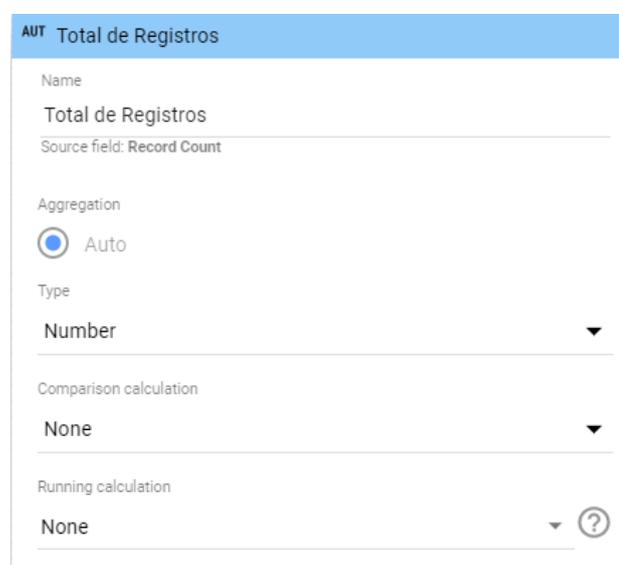


Figura 4.21. Edição do nome de métrica.

4.5.3.5. Available Fields

Nesta área, todas as colunas da base de dados importada são convertidas em campos, que podem ser modelados como dimensões ou métricas a serem analisadas. Estes campos serão a base para gerar funções, operadores e inserir a lógica nos dados. Cada campo possui uma tipagem diferente, podendo ser caracterizadas como strings, booleanos ou valores numéricos, por exemplo.

4.5.4. Configurando e Gerando Novas Visualizações

Ao configurar o *Scorecard*, é possível observar que a dimensão é categorizada como o campo “Dia”, pois a data possui dados únicos para agregar os registros. A métrica em si será o campo numérico “*Record Count*”, gerado automaticamente pelo Google Data Studio para contabilizar o valor geral de todos os registros presentes na importação. Por fim, será necessário editar o nome da métrica, substituindo para “Total de Registros”.

Partindo para outra visualização, desta vez será mostrada a “proporção racial das mulheres”. Para ilustrar melhor esses dados, será selecionada a opção do Gráfico de Pizza, localizado conforme a Figura 4.22.

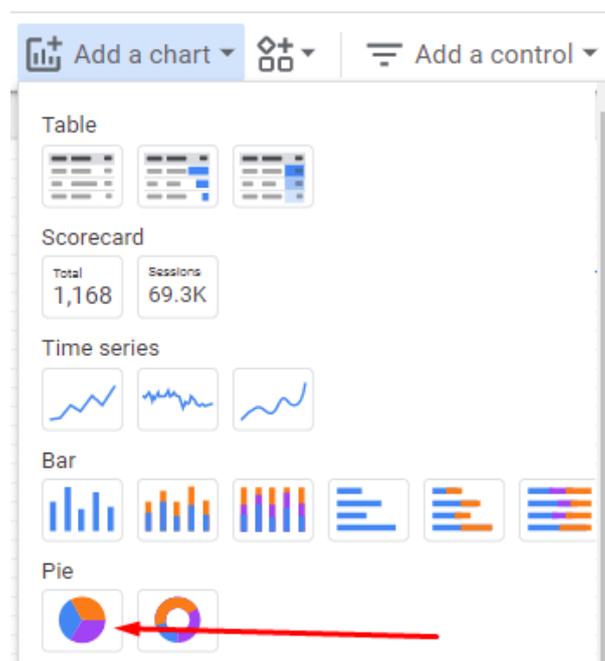


Figura 4.22. Opção da visualização de gráficos de setores ou de pizza (*Pie chart*).

Para configurar o gráfico, é preciso mudar a dimensão padrão que está sendo utilizada. Assim, a dimensão atual deverá ser alterada para o campo “cor”, que representa as raças reportadas. A métrica será o total de registros, pois será necessária a proporção de respostas de acordo com a quantidade de itens especificados na dimensão. O resultado final pode ser visto na Figura 4.23.

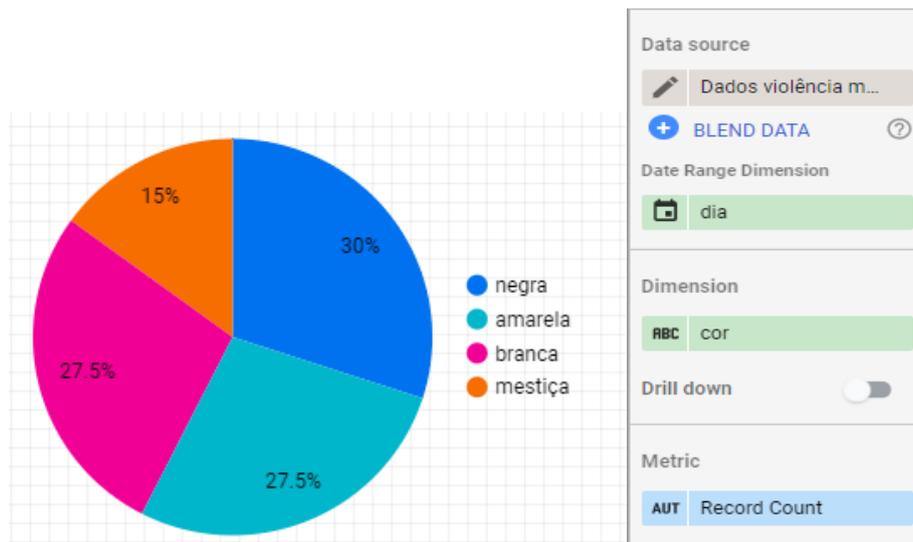


Figura 4.23. Gráfico de pizza gerado, seguido de suas propriedades.

Para finalizar as visualizações, será gerado um gráfico para mostrar a relação do agressor com a vítima. Para ilustrar melhor esses dados, será selecionada a opção do Gráfico de barras em colunas, localizado conforme a Figura 4.24.

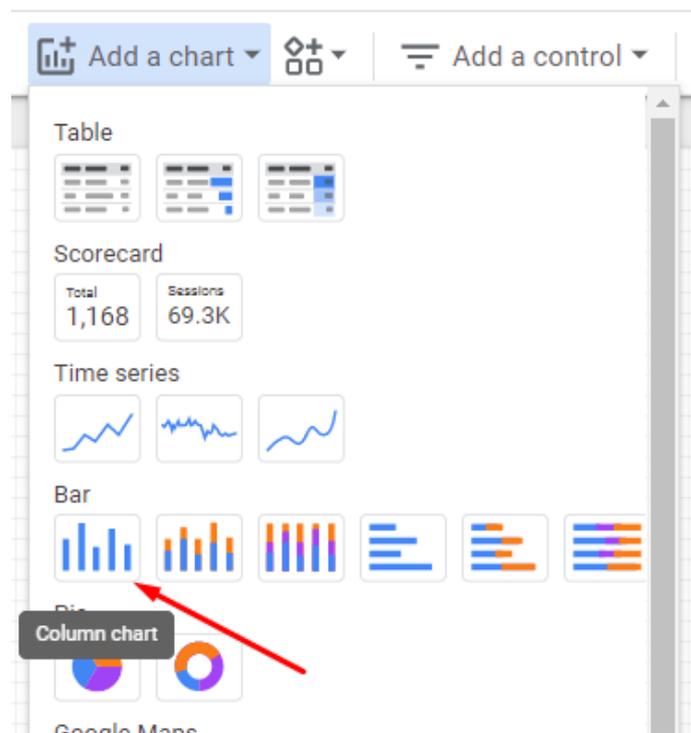


Figura 4.24. Opção da visualização de gráficos de barras em colunas.

Para configurar o gráfico, é preciso novamente mudar a dimensão padrão que está sendo utilizada. Assim, a dimensão atual é alterada para o campo “rel_agressor”, que representa o grau de parentesco do agressor. A métrica continuará sendo o total de registros. O resultado final pode ser visto na Figura 4.25.

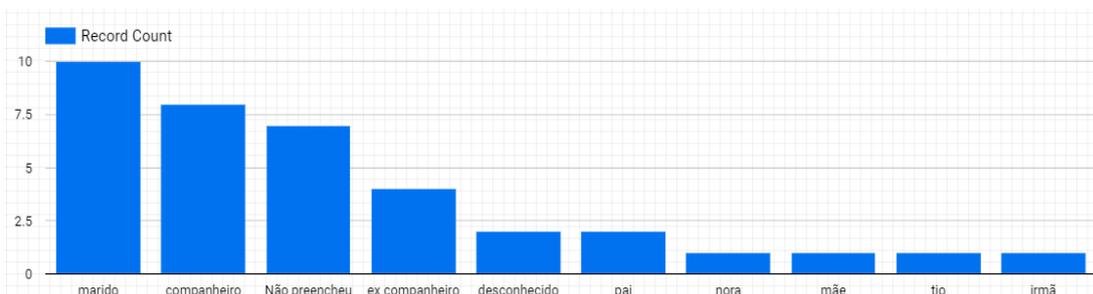


Figura 4.25. Gráfico de colunas gerado, seguido de suas propriedades.

Caso haja a necessidade de mudar a cor do gráfico de barras, é possível mudar dentro da aba lateral de propriedades da visualização. Na aba *Style*, será mudada a cor das barras de azul para vermelho, conforme Figura 4.26.

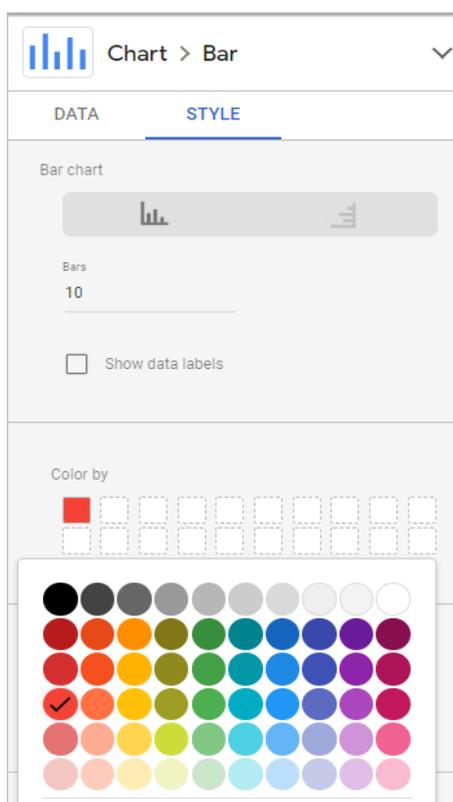


Figura 4.26. Gráfico de colunas gerado, seguido de suas propriedades.

A página é então finalizada com três visualizações principais: A quantidade de registros, a porcentagem de raças informadas e a comparação entre os agressores reportados.

4.5.5. Decisões de Design

Dado que a visualização gerada trabalha com dados sensíveis, é extremamente importante garantir que as informações sejam transmitidas de maneira que se consiga capturar a atenção e despertar os sentimentos corretos das pessoas que irão interagir com ela. Dessa forma, a implementação da identidade visual foi concebida através de um conjunto de regras de design durante a montagem do *dashboard*.

4.5.5.1. Escolha das cores

Uma das técnicas implementadas se dá pelo uso correto das cores a serem implementadas no *dashboard*. Goethe (1967) iniciou os primeiros estudos voltados ao efeito da influência das cores na percepção do ser humano. A área atual, denominada de psicologia das cores, aborda um conjunto de sentimentos e sensações que determinada cor pode transparecer, de acordo com o contexto em que ela está sendo abordada.

Considerando o contexto de violência e urgência explicitados nos dados, foram escolhidas três cores principais para compor o *dashboard*: a cor preta, compondo o sentimento de pesar e luto, a cor vermelha, compondo o sentimento de perigo e urgência e a cor branca, simbolizando traços de simplicidade e honestidade dos dados. A paleta de cores escolhida está dentro dos conformes de acessibilidade de acordo com Web Content Accessibility Guidelines (WCAG) [Caldwell, Cooper, Reid, Gregg *et al.* 2008], garantindo também a facilidade na leitura dos dados através de dispositivos de diferentes resoluções.

4.5.5.2. Proporção entre as cores

Com a paleta de cores escolhida, também foi necessário definir onde as mesmas seriam aplicadas. Uma das técnicas principais utilizadas para facilitar essa tomada de decisão foi o princípio 60-30-10, que divide uma paleta de três cores a serem utilizadas de acordo com essa proporção no projeto. A cor primária, compondo aproximadamente 60% da composição, foi a cor preta, representando o fundo utilizado na interface. Já alguns detalhes e gráficos foram montados com a cor secundária vermelha, representando 30% da composição na interface. Por último, foi utilizada a cor branca para destacar alguns detalhes e legendas nos gráficos montados, representando 10% da composição.

4.5.6. Estilizando a Página

Seguindo o princípio do 60-30-10, será criado um fundo preto para o nosso *dashboard*. Sem selecionar nenhum gráfico, o menu da lateral direita do Data Studio irá mostrar algumas opções de Temas e *Layout* customizados. Na aba “Theme”, clique no botão *Customize* e troque a cor do fundo em “*Report Background*” para preto. O resultado pode ser visto conforme a Figura 4.27.

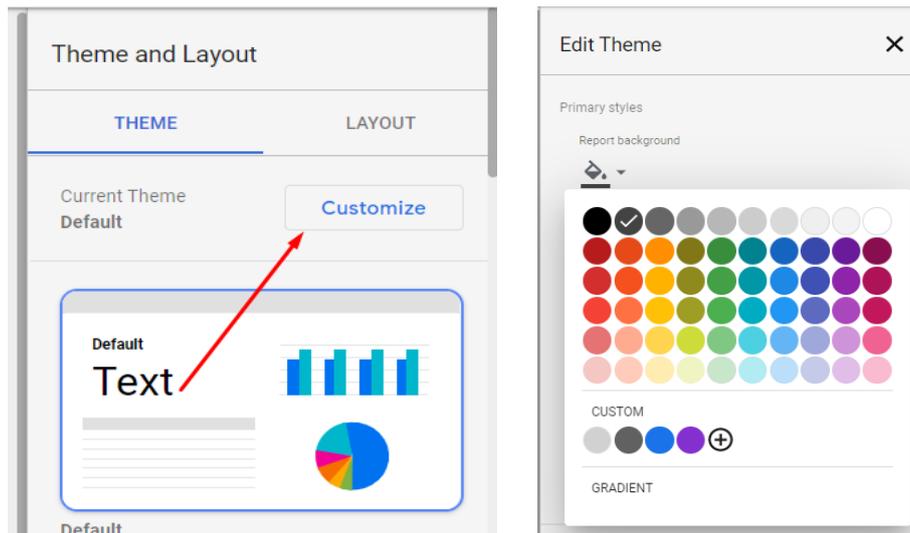


Figura 4.27. Customização de cor base do *dashboard*.

É possível também alterar a fonte e a cor geral dos textos. Na mesma aba onde foi alterado o fundo, mude a fonte para “Oswald” e a cor para branco, conforme a Figura 4.28.

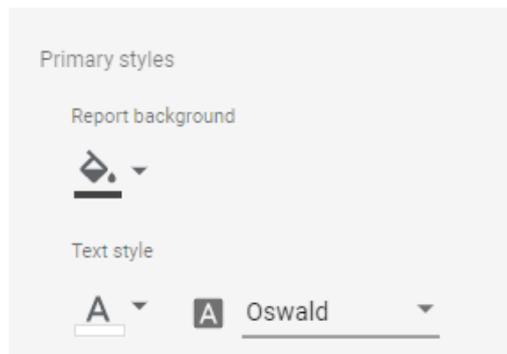


Figura 4.28. Customização de texto base do *dashboard*.

O próximo passo é gerar algumas divisórias e textos para os gráficos em si. Criando formas geométricas, é possível usar isso como uma espécie de container para as visualizações.

Crie um retângulo clicando no ícone no *menu* principal, e insira as devidas propriedades para dar fundo, borda e cor à forma, conforme a Figura 4.29.

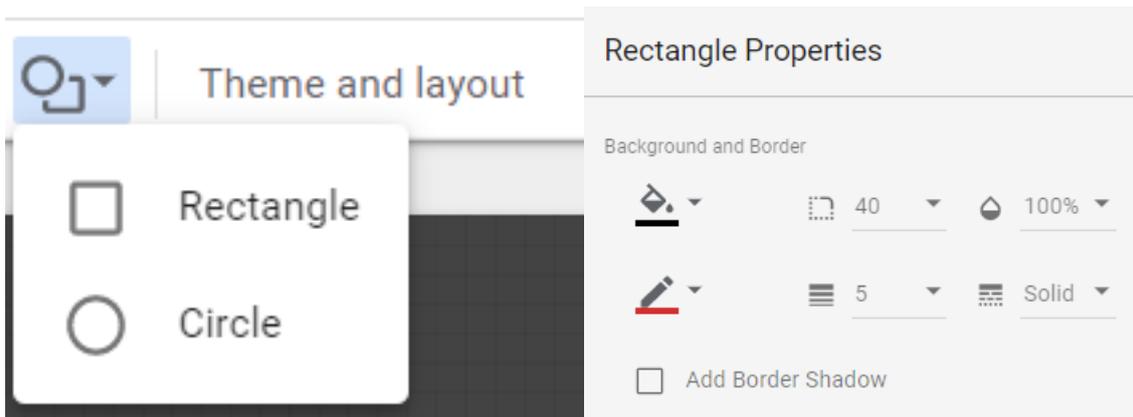


Figura 4.29. Criando formas e alterando as suas propriedades de borda e cor.

É possível então inserir uma caixa de texto dentro das nossas divisórias, de modo a ilustrar melhor as visualizações em si. Um título também será inserido do lado de fora, para representar o nome da pesquisa. Clique no campo de inserir Texto e insira os textos conforme a Figura 4.30 e ajuste-os para chegar em algo similar à Figura 4.31.

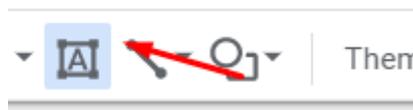


Figura 4.30. Criando textos.

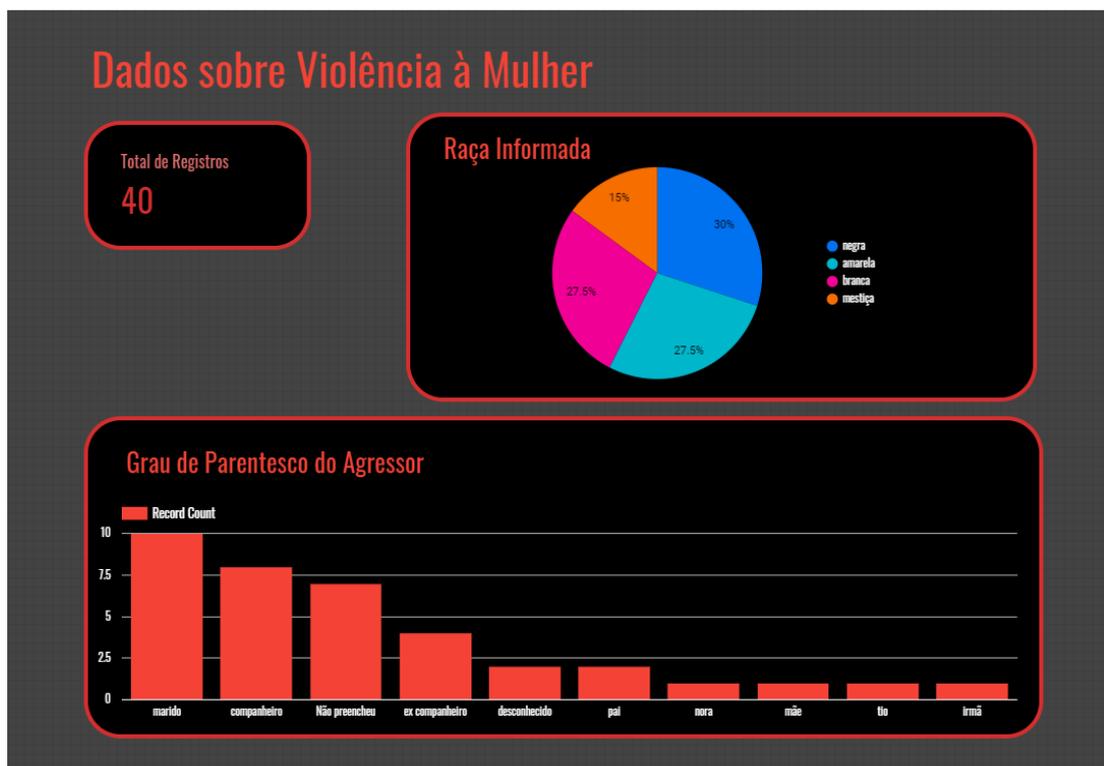


Figura 4.31. Produto final: a visualização gerada.

4.6. Considerações Finais

Neste capítulo foram abordadas técnicas e práticas de Jornalismo de Dados tais como coleta de dados, anonimização de bases de dados, limpeza e transformação de dados utilizando um banco de dados MySQL como ferramenta e sumarização de informações por meio de um *dashboard* interativo. Para isto foi utilizado como evento propulsor, um caso real de jornalismo de dados no contexto da violência contra a mulher, que se manifestou durante uma pesquisa realizada pelo grupo de literacia de dados do PPGI da UFRJ com dados sigilosos do Centro Especializado de Atendimento à Mulher de um município do Estado do Rio de Janeiro.

Espera-se que este material seja útil não somente para divulgar conhecimento sobre os recursos e ferramentas disponíveis para a atividade jornalística por meio de dados, como também para melhorar competências e atitudes relacionadas à literacia de dados de estudantes e pesquisadores brasileiros, sensibilizando-os para a temática da violência contra a mulher e abrindo portas para uma relação ainda mais integrada dos cidadãos com uma das demandas sociais mais urgentes da humanidade, que é o combate à violência contra a mulher.

Referências

- ABRAJI (2012) “Jornalismo Investigativo - definições de associados e seguidores”. Disponível em: <https://abraji.org.br/noticias/jornalismo-investigativo-definicoes-de-associados-e-seguidores>. Acesso em agosto de 2021.
- Alves, M. C. L., Dumaresq, M. L. and Silva, R. V. (2016) “As Lacunas no Enfrentamento à Violência contra a Mulher: análise dos bancos de dados existentes acerca da vigilância doméstica e familiar”. Brasília: Núcleo de Estudos e Pesquisas Senado. Disponível em <https://www2.senado.leg.br/bdsf/handle/id/519161>. Acesso em março de 2021.
- Appelgren, E. and Nygren G. (2014) “Data journalism in Sweden: Introducing new methods and genres of journalism into ‘old’ organizations”, *Digital journalism*, v. 2, n. 3, p. 394-405.
- Bazzi, C. L. (2013) “Introdução a Banco de Dados”. Editora UFTPR, Curitiba.
- Bioni, B. R. (2019) “Proteção de dados pessoais: a função e os limites do consentimento” Rio de Janeiro: Forense.
- Borburema, T. L. R., Pacheco, A. P., Nunes, A. A., Moré, C. L. O. O. and Krenkel S. (2017). “Violência contra mulher em contexto de vulnerabilidade social na Atenção Primária: registro de violência em prontuários”. *Revista Brasileira de Medicina de Família e Comunidade*.;12(39):1-13. [http://dx.doi.org/10.5712/rbmfc12\(39\)1460](http://dx.doi.org/10.5712/rbmfc12(39)1460).
- Bounegru, L., Chambers, L. and Gray, J. (2012) “The Data Journalism Handbook 1”, Produced by European Journalism Centre, Disponível em: <https://datajournalism.com/read/handbook/one>. Acesso em Agosto de 2021.
- Brasil. (2006) “Lei Maria da Penha”. Lei nº. 11.340, de 7 de agosto de 2006.
- Brasil. (2018) “Lei Geral de Proteção de Dados Pessoais”. Lei nº13.709, de 14 de agosto de 2018.

- Burelomova, A. S., Gulina, M. A. and Tikhomandritskaya, O. A. (2018) “Intimate partner violence: An overview of the existing theories, conceptual frameworks, and definitions”, *Psychology in Russia: State of the art*, v. 11, n. 3, p. 128-144.
- Caldwell, B., Cooper, M., Reid, L. G., Gregg Vanderheiden *et al.* (2008) *Web Content Accessibility Guidelines (WCAG) 2.0*. W3C Recommendation 11 December 2008. Disponível em <https://www.w3.org/TR/WCAG20/>. Acesso em Maio de 2021.
- Ferreira F. (2012) “Afimal, o que é o jornalismo?” Disponível em: <http://www.observatoriodaimprensa.com.br/feitos-desfeitas/ed719-afinal-o-que-e-jornalismo/>. Acesso em agosto 2021.
- Goethe W. (1810/1967). “Theory of Colours”. London: Frank Cass.
- GovLab (2021) “Data Collaboratives”. Disponível em <https://datacollaboratives.org/>. Acesso em agosto de 2021.
- Michaelis (2021) “Dicionário Brasileiro da Língua Portuguesa” Disponível em <https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=an%C3%B4nimo>. Acesso em abril de 2021.
- Open Knowledge Foundation (2021) “Open Definition”. Disponível em <http://opendefinition.org/>. Acesso em agosto de 2021.
- Oracle (2021) “Banco de dados definido” Disponível em: <https://www.oracle.com/br/database/what-is-database/>. Acesso em agosto de 2021.
- Oracle Corporation and/or Its Affiliates (2021) “Why MySQL?” Disponível em: <https://www.mysql.com/why-mysql/>. Acesso em agosto de 2021.
- Villars, R. L., Olofson, C. W. and Eastwood, M. (2011) “Big data: What it is and why you should care”, White paper, IDC, v. 14, p. 1-14.
- World Health Organization *et al.* (2020) “COVID-19 and violence against women: what the health sector/system can do”, 7 April 2020. World Health Organization.



Luciana Sá Brito é Doutoranda em Informática/Gestão de Sistemas Complexos (UFRJ), Mestre em Informática (UFRJ/2020), Especialista em EAD/Educação Profissional (SENAC/2012), Licenciada em Física (UFRJ/2008) e Técnica em Artes Dramáticas (ETE Martins Pena/2010). Luciana atua como Designer Instrucional na Fundação CECIERJ desde 2014 e foi Física Docente na SEEDUC-RJ (2007-2014). Seus interesses principais estão relacionados à pesquisa em Design Instrucional e Sistemas Colaborativos, pesquisa em Interação Humano Computador e à Ciência de Dados como forma de ativismo. Nas horas livres pratica montanhismo e toca piano.



Alayne Duarte Amorim é Doutoranda em Computação pela UFRJ e Mestre em Informática pelo UFRJ/2012. Possui especialização em Gerência de Tecnologia de Computação pela UFF/2003 e especialização em Administração de Banco de Dados pela Universidade Estácio de Sá em 2006. Licenciada para disciplinas de informática para o ensino fundamental e médio pelo Instituto a Vez do Mestre (2009) e graduada em SI pela Universidade do Grande Rio (1999). É professora e coordenadora pedagógica do curso técnico integrado em Informática no Colégio Pedro II.



André Viana Tardelli é Mestrando em Informática pelo grupo GRECO/PPGI da Universidade Federal do Rio de Janeiro e coordenador dos cursos de User Experience (UX) E Design no Grupo Alura. Também atua como instrutor na Caelum, ministrando aulas dos cursos de UX, Desenvolvimento Web e de Python voltado para Ciência da Dados.



Angélica Fonseca da Silva Dias é Doutora em Informática pelo PPGI/UFRJ. Atualmente é Diretora do Instituto Tércio Pacitti-NCE/UFRJ, Docente colaborador da UFRJ e do SEMESP/MEC. Coordenadora do Lab. De Pesq. Computacionais em Economia Circular voltados para os ODS-Agenda 2030. Membro Consultivo da Comissão Especial de Sistemas Colaborativos (CE-SC) – SBC. Tem experiência na área de Sistemas de Informação. Mãe de duas filhas lindas. Ama ler, pilates e uma boa corrida.



Juliana Baptista dos Santos França é Doutora em Informática pelo PPGI/UFRJ (2018) e concluiu seu Pós-doc no PPGI/UFRJ em 2019. Atualmente é docente no IC/UFRJ. Atua no programa de pós-graduação de Gestão e Estratégia (PPGE/UFRRJ) como membro permanente desde 2021 e colabora com o PPGI/UFRJ desde 2018. É membro da Comissão Especial de Sistemas Colaborativos (CE-SC) da SBC desde 2019 e também da ACM/FCA no mandato de 2019-2021. Seu principal interesse está na área de CSCW, com contribuições aos domínios de Suporte à Decisão, Gestão por Processos de Negócio, Crises e Desastres, e Informática na Educação, mas não limitado a estes.



Adriana Santarosa Vivacqua é Professora do Instituto de Computação da UFRJ. Obteve o doutorado na COPPE/UFRJ em (2007). Foi bolsista de Produtividade do CNPq e Jovem Cientista da FAPERJ. Seus interesses de pesquisa incluem IHC inteligente, CSCW, visualização. Atua como Chair Associado para Equidade junto ao Comitê Executivo da ACM SIGCHI, e foi eleita recentemente para a posição de VP-at-Large no Comitê Executivo da ACM SIGCHI (mandato 2021-2024).

Capítulo

5

ENDO-GDC: Projetando Jogos Educacionais

Geraldo Xexéo e Bernardo Taucei

Abstract

This tutorial presents a general introduction to games and game design for educational purposes, followed by step-by-step execution of a conceptual design, with theoretical foundation, using ENDO-GDC, a canvas for the Design of Endogenous Educational Games, where content and rules are integrated.

Resumo

Este tutorial apresenta uma introdução geral a jogos e projetos de jogos com fins educacionais, seguida do detalhamento passo a passo, com fundamentação teórica, do ENDO-GDC, um canvas destinado ao Design de Jogos Educacionais Endógenos, onde conteúdo e regras são integrados.



Figura 5.1. O canvas de projeto de jogos educacionais ENDO-GDC. Fonte: própria

5.1. Introdução

Este capítulo é uma introdução ao uso do *Endogenous Game Design Canvas* (ENDO-GDC) [Taucei, 2019], apresentado na Figura 5.1, um canvas destinado a criação de propostas conceituais de jogos educacionais endógeno. Um jogo educacional endógeno, como se verá mais adiante, é um jogo onde o conteúdo educacional, ou a mensagem, está imbricado nas suas regras e mecânicas.

O ENDO-GDC pode ser empregado no design ou na análise tanto de jogos digitais quanto o de jogos não-digitais, como jogos de tabuleiro, de carta e até jogos assemelhados a esporte. Os conceitos utilizados são genéricos e podem ser especializados para qualquer tipo de jogo. Entre os benefícios proporcionados pelo uso do ENDO-GDC estão:

- ser de simples entendimento;
- facilitar o alcance de acordos comuns, especialmente para grupos multi-disciplinares;
- ter uma base teórica sólida, e
- também poder ser usado para a análise e melhoria de jogos existentes.

Neste tutorial, junto com a apresentação do ENDO-GDC, será também explicado como iniciar o processo de criação de jogos educacionais destinados a contextos variados de ensino formal, informal, ou semi-formal[CEDEFOP, 2014]. Mas antes do tutorial propriamente dito, é feita uma introdução à teoria de design de jogos e outras teorias relevantes ligadas a jogos educacionais e sérios. Essa introdução será suficiente para permitir ao leitor conhecer a justificativa para cada passo do ENDO-GDC. O leitor interessado em se aprofundar mais nesses tópicos encontrará nas referências bibliográficas uma rica variedade de fontes

5.2. Motivação

Muitos estudantes e profissionais de Sistemas de Informação e carreiras afins veem o desenvolvimento de jogos como uma carreira atrativa. Além de existir uma motivação intrínseca para participar do desenvolvimento de algo que os fascinou desde a infância e a adolescência, há também uma motivação econômica: o mercado internacional de jogos alcançou US\$175.8 bilhões no ano de 2021 [Newzoo, 2021], garantindo a existência de empregos e oportunidades de empreendedorismo na área. Ainda pequeno em relação ao mercado de jogos em geral, o de jogos educacionais já atingiu a marca de US\$9.2 bilhões em 2019 e deve seguir crescendo a US\$88.11 bilhões até 2027, abrindo mais um leque de oportunidades[Verified Market Research, 2020]

Entre os fatores primários desse crescimento estão: a necessidade de um maior engajamento dos usuários entre empreendimentos, o desenvolvimento no uso de jogos educacionais baseados em dispositivos móveis e o progresso nos resultados de aprendizado [Verified Market Research, 2020]. Entretanto, o *game design* inadequado e a falta de proficiência em jogos sérios previne o crescimento do mercado [Verified Market Research, 2020].

Por que precisamos de jogos educacionais? Há um consenso entre pesquisadores da área de jogos na educação que **jogos são fortemente motivadores**, sendo vistos como um importante aliado no ensino-aprendizagem [Prensky, 2001; Prensky, 2003; L. D. Costa, 2008; McGonigal, 2011; Kalmpourtzis, 2018]. Mesmo jogos cujo obje-

tivo central é o entretenimento podem motivar o jogador a aprender [Koster, 2004]. Um exemplo conhecido é o estudo voluntário de probabilidade por jogadores de *fantasy baseball* nos EUA, que motiva inclusive um curso na PBS¹. Outro exemplo é o conhecimento geográfico básico proporcionado pelo jogo “War”.

Ao mesmo tempo, há outro consenso: que a educação tradicional, criticada como a escola do “cuspe e giz”, está fortemente ameaçada pelas mudanças da sociedade, principalmente as causadas pelas novas tecnologias da comunicação e da informação. Técnicas como a sala de aula invertida [Bishop e Verleger, 2013], Aprendizado Baseado em Jogos [Plass, Meyer e Homer, 2019], Aprendizado Baseado em Projetos [Larmer, Mergendoller e Boss, 2015], e outras metodologias ativas buscam tornar a sala de aula, considerada aborrecida, mais atrativa ao aluno, aumentando seu engajamento, facilitando o aprendizado e tornando aluno protagonista do processo de aprendizagem. A presença dos jogos em sala de aula ou em ambientes on-line faz parte de várias propostas de modernização do ensino-aprendizagem [Kalmpourtzis, 2018].

Vários autores também defendem técnicas específicas de ensino-aprendizagem usando jogos, sendo que o termo Aprendizado Baseado em Jogos, ou *Game Based Learning* (GBL), pode ser usado amplamente para descrever várias formas de uso dos jogos para o ensino-aprendizagem, incluindo diferentes arcabouços [Plass, Meyer e Homer, 2019], jogos educacionais específicos ou mesmo o uso de jogos de entretenimento em sala de aula [Prado, 2018].

Van Eck [2006] expõe como os jogos educacionais incorporam todas as fases que ativam processos necessários para uma aprendizagem efetiva, presentes nos nove eventos de Gagne et al. [2004]: ganhar atenção, informar os objetivos do aluno, estimular a recordação da aprendizagem anterior, apresentar material de estímulo, fornecer orientação ao aluno, provocar o desempenho, fornecer *feedback*, avaliar o desempenho e aprimorar a retenção e a transferência.

Já Lieberman [2006] lista alguns benefícios do uso de jogos para aprendizagem:

- eles proporcionam ao jogador uma experiência ativa;
- encorajam o jogador a aprender fazendo;
- são um meio social que proporciona ao jogador interações entre pessoas e respostas emocionais;
- são participativos, fornecendo ao jogador *feedback* personalizado e rápido;
- são envolventes. A participação faz com que o jogador preste muita atenção. Requer um planejamento e tomada de decisão cuidadosa. Exige aprendizagem para ter sucesso;
- promovem a aprendizagem comportamental, dando ao jogador recompensas por comportamento (pontos, poder, classificação e assim por diante), e esse *feedback* positivo no jogo pode encorajar comportamentos desejados no mundo real;
- oferecem consequências, que não são abstratos ou hipotéticos, mas representadas diretamente. O jogador interpreta um personagem e se identifica com ele ou ela. Sucesso e fracasso mapeiam diretamente as suas ações; o ego e a auto-imagem são

¹Public Broadcasting Service, é uma rede de televisão americana de carácter educativo-cultural, sem publicidade.

investidos na experiência, e

- eles fornecem modelos para o jogador que pode aprender com os personagens do jogo e entender suas experiências comportamentais.

Além disso, Susi, Johannesson e Backlund [2007] e Squire e Jenkins [2003] mostram como jogos educacionais possibilitam a vivência de situações que não podem ser experimentadas no mundo real por razões de segurança, custo, tempo, etc.

Diversas revisões sistemáticas, nos últimos anos, têm tratado o tema *Game Based Learning*. Abdul Jabbar e Felicia [2015] concluem que GBL “ajuda os estudantes a desenvolver habilidades [skills] e conhecimento e fortalece sua habilidade de tratar as experiências de aprendizado fornecida pelos jogos”. Já Tsutsumi et al. [2020] encontram 24 estudos que mostram que os “os jogos tiveram o efeito de aumentar o desempenho acadêmico de alunos”. Mais recentemente, S. B. Júnior [2020] selecionou 180 estudos primários entre 2009 e 2020 concluindo “que os jogos têm sido cada vez mais utilizados como uma ferramenta adicional ao ensino e aprendizagem, possibilitando desenvolver o conhecimento e habilidades do público-alvo”.

Possivelmente, há um viés positivo nos relatos publicados, já que não há o costume de publicar resultados negativos, porém, devido ao acúmulo de evidências, acreditamos ser possível considerar que jogos, e o aprendizado baseado em jogos, trazem benefícios para o ensino-aprendizagem. Marklund [2015], apesar de apoiar o uso de jogos na educação, diz que seu uso na educação formal “introduz demandas pesadas para a infraestrutura, cultura e processos de trabalho de uma instituição”, fatores pouco citados pela maioria dos artigos no assunto, que, quando fazem uma avaliação de um jogo ou de um método, o fazem de forma isolada e não no contexto de introdução de mudanças em uma escola ou universidade.

Dessa forma, os jogos educacionais oferecem oportunidades para melhoria do aprendizado, mas apenas se o jogo for bem projetado [Winn, 2009; Kalmpourtzis, 2018; Verified Market Research, 2020]. A pergunta em aberto é: como fazê-lo? Não seria a mesma coisa que desenvolver jogos de entretenimento?

Respondendo essa pergunta, a mesma pesquisa de Marklund [2015], para sua tese de doutorado, concluiu que “os processos associados com o desenvolvimento de jogos ‘tradicionais’ são incomensuráveis com o desenvolvimento de jogos educacionais”.

Para Winn [2009] fazer um bom jogo é difícil porque o processo de design de jogos, além de criativo, é às vezes é caótico. Todavia, desenvolver um jogo educacional é uma tarefa ainda mais difícil, porque em vez de simplesmente tentar otimizar o aspecto de entretenimento do jogo, também é preciso otimizá-lo para alcançar um conjunto específico de resultados de aprendizado.

Fica claro que o caráter educacional traz desafios adicionais à já difícil tarefa de desenvolver jogos. Além de requerer equipes ainda mais multi-disciplinares, o que acentua as dificuldades de comunicação no projeto, os requisitos fundamentais do desenvolvimento de um jogo, como a necessidade de provocar emoções e entretenimento, podem se chocar com as necessidades educacionais [L. D. Costa, 2008]. Isso significa que são necessárias estratégias específicas para a criação de jogos educacionais. O emprego de canvas aparece como uma alternativa para se mitigar tanto os riscos de comunicação

quanto o conflito de requisitos.

5.2.1. Motivação para o uso de *canvas*

O *canvas* é um quadro ou painel, usado em uma reunião, que ajuda a sintetizar e organizar as informações sobre o artefato que se deseja controlar ou desenvolver, como, por exemplo, um modelo de negócio [Osterwalder e Pigneur, 2010]. Uma característica importante dos *canvas* é serem de fácil compreensão para toda a equipe. Por isso, seu uso é capaz de proporcionar uma base comum de pensamento entre todos os participantes, e um acordo transparente e de alto nível acerca do que deve ser feito ou atingido.

Normalmente um *canvas* é caracterizado por dois tipos de objetos: um quadro grande na parede dividido em áreas com significados específicos, e *post-its* usados para registrar informações e acordos feitos dentro dessas áreas. Versões digitais podem ser facilmente construídas usando software de apresentação, como o Google Slides, ou espaços compartilhados de desenho, como o Miro. Existem ainda ferramentas dedicadas para um ou vários tipos de *canvas*, como o Canvanizer.

O divisão do quadro em áreas indica que assuntos devem ser tratados. Normalmente essas áreas devem ser preenchidas em uma ordem, e possuem associações com as áreas próximas a elas, ou em um mesmo bloco. Assim, cada *canvas* é usualmente apresentado junto com um método de preenchimento, ou de execução da reunião.

O uso de *post-its*, ou sua versão digital, permite que as anotações, ou registros, possam ser vistas, discutidas e alteradas com facilidade, sendo uma prática comum nos métodos ágeis de gestão, adequados ao desenvolvimento de jogos, principalmente por serem normalmente iterativos e incrementais.

Já existem algumas iniciativas de *canvas* específicos para o design de jogos, revisados por Taucei, L. Costa e G. Xexéo [2018] e Taucei [2019], como as propostas de Carey [2015] e Lam [2015], porém, não encontramos nenhuma iniciativa destinada especificamente a criação de jogos educacionais.

Detectamos então, em 2017, a oportunidade de desenvolver um *canvas* para a criação de jogos educacionais, tendo em vista os relatos positivos de uso de *canvas* em várias indústrias, inclusive a de jogos.

Essa é a missão do *Endogeneous Game Design Canvas* (ENDO-GDC) [Taucei, 2019]. Com ele é possível criar e manter o design conceitual de um jogo educacional, de modo a constituir uma base conceitual comum entre educadores e desenvolvedores. O uso do ENDO-GDC facilita a compreensão de como o jogo funciona e como atinge seus objetivos educacionais.

5.3. O que são jogos

Como alternativa ao referencial teórico apresentado nesta seção, existe uma definição bem mais simples do que é um jogo: **jogo é aquilo que você joga**. Essa definição lembra a Teoria Ingênua de Conjuntos, que não define os termos conjunto, elemento e pertinência. Se esta definição existe, por que é necessário definir jogos, se todos podem reconhecer o que é um jogo? A resposta é: porque precisamos entender profundamente o artefato e a experiência que queremos criar.

Para construir um jogo é necessário adotar uma visão do que é um jogo. Isso é uma tarefa reconhecidamente difícil porque a palavra jogo tem muitos significados, principalmente em outras línguas que não o português. O francês *jouet*, o inglês *to play* e o alemão *spiel*, por exemplo, possuem traduções diferenciadas para o português como jogar, atuar e brincar.

Mesmo em português, a palavra jogo possui vários outros significados, como:

- o jogo₁ como objeto, como o tabuleiro e peças de xadrez;
- o jogo₂ como partida registrada, como uma gravação, um relato ou por meio de uma notação;
- o jogo₃ como a definição abstrata, composta de regras e que viabiliza uma infinidade de partidas possíveis, e
- o jogo₄ como atividade, que se desenrola entre jogadores, dentro de um contexto, em um momento no tempo e um lugar no espaço, e que usa os objetos, as regras abstratas e pode ser registrada como um jogo-partida.

Assim, uma pessoa compra um jogo₁ de Xadrez, para jogar₄ Xadrez com seus amigos no fim da aula, e guarda os jogos₂ para estudá-los mais tarde. Tudo isso seguindo as regras do jogo₃ Xadrez. O mesmo jogo₁ pode ser usado para jogar outro jogo₃, uma variante conhecida como Xadrez 960.

Em vista de tantas possibilidades, este tutorial a definição de jogo proposta por G. Xexéo, Carmo et al. [2017]:

Jogos são atividades sociais e culturais voluntárias, significativas, fortemente absorventes, não-produtivas, que se utilizam de um mundo abstrato, com efeitos negociados no mundo real, e cujo desenvolvimento e resultado final é incerto, onde um ou mais jogadores, ou equipes de jogadores, modificam iterativamente e de forma quantificável o estado de um sistema artificial, possivelmente em busca de objetivos conflitantes, por meio de decisões e ações, algumas com a capacidade de atrapalhar o adversário, sendo todo o processo regulado, orientado e limitado, por regras aceitas, e obtendo, com isso, uma recompensa psicológica, normalmente na forma de diversão, entretenimento, ou sensação de vitória sobre um adversário ou desafio.

Esta é uma definição bastante detalhada, criada a luz de várias outras pré-existentes, como, por exemplo as de Huizinga [1938–2001], Salen e Zimmerman [2003], Callois [2001], Juul [2012] e Suits e Hurka [2014], e se destina a marcar os jogos como objeto de estudo, especificamente os jogos de entretenimento, dentro de uma área ampla que se autodenomina de várias formas, como *game design* ou *game studies*. O leitor interessado em se aprofundar em cada conceito apresentado pode procurar fontes adicionais a partir de G. Xexéo, Carmo et al. [2017].

Neste tutorial, o interesse é no jogo como uma atividade, jogo₄, que acontece em algum contexto [Huizinga, 1938–2001; Järvinen, 2008]. Essa atividade é significativa por si só [Salen e Zimmerman, 2003], o que é uma afirmação importante, pois contradiz a visão de alguns autores que estudaram o jogo com um significado outro, como servir de preparação das crianças para vida real [Sutton-Smith, 1997].

O contexto é formado pelos jogadores, em seus estados cognitivos e emocionais, e por uma miríade de fatores que colocam o jogo em um intervalo de tempo, lugar, com participantes e plateia, sob diferentes formas de pressão, etc [Järvinen, 2008].

Como atividade, o jogo só se realiza quando tirado da caixa, ou ligado no computador, console ou celular, e jogado. A partir do início do jogo, se apresenta uma quantidade, normalmente enorme, e potencialmente infinita, de partidas possíveis, algumas possivelmente indiferenciáveis na prática. Naquele intervalo de tempo, porém, só uma dessas partidas se realizará. Os jogadores, interagindo entre si e com o jogo₃ e usando o jogo₁, escolhem um caminho, criando a partida, o jogo₂, e recriando o jogo₃, inclusive com a possibilidade de adotar regras próprias nessa jornada. O jogo não pertence mais ao designer, mas, agora, pertence aos jogadores.

Duas partidas nas quais as mesmas jogadas acontecem podem proporcionar experiências totalmente diferentes, dependendo do que acontece ao nível de contexto, incluindo aí os jogadores, o ambiente, a comunicação entre eles e todos os outros fatores associados à atividade de jogar. Por exemplo, em duas partidas nas quais exatamente os mesmos movimentos foram executados, dois jogadores de xadrez podem um ensinar ao outro, ou um provocar o outro, trazendo emoções diferentes para jogadores e plateia, a partir de dois jogos₂ que serão posteriormente vistos como iguais quando analisados de forma abstrata, a partir do registro da partida na notação adequada.

A característica mais importante de jogos é que possuem regras. São elas que permitem que os jogadores joguem. Aceitar as regras arbitrárias para poder fruir da experiência do jogar é conhecido como “atitude lusória” [Suits e Hurka, 2014]. As regras são um grande diferencial dos jogos para outras atividades de entretenimento, como livros e cinema [Crawford, 2003], ou do livre brincar [Callois, 2001].

Também essencial aos jogos é a agência do jogador. É o jogador que faz o jogo acontecer, e ele cria caminhos para o desenrolar do jogo, fazendo com que algo que é apenas potencial aconteça. Ele então modifica constantemente o estado do jogo, com o intuito de alcançar um objetivo, normalmente ganhar o jogo sozinho, mas cada vez mais, devido à influência dos RPGs² e dos videogames, ganhar **do** jogo, com seus companheiros, nos chamados “jogos colaborativos”.

Esse entendimento do jogo permite ver o jogo como um sistema [Salen e Zimmerman, 2003], que passa por estados, sendo basicamente um sistema de informações [G. Xexéo, Eduardo Mangeli et al., 2021], o que pode ser muito útil na criação de jogos, pois além de nos fornecer uma maneira de pensar o jogo, também permite estudar formalmente suas propriedades.

Quando se associa um novo objetivo a um jogo, como educar, convencer ou fazer propaganda, um novo significado é também associado a ele, que não é apenas aquele intrínseco ao jogar e aos significados culturais do jogo ou do jogar. De certo modo, quebramos o Círculo Mágico [Huizinga, 1938–2001], que isola o jogador do mundo real

²*Role-Playing Games* são jogos onde um mestre do jogo guia os jogadores por aventuras imaginárias, onde os jogadores assumem o papel de personagens. O pioneiro e provavelmente mais conhecido é o *Dungeons & Dragons*, onde as aventuras acontecem em um mundo medieval com magias, monstros e raças.

ao seu redor, transformando-o em uma película permeável, por onde passam, intenções, contextos e situações.

Outra característica importante de ressaltar, é que o ato de jogar é uma atividade cognitiva, onde o jogador observa uma situação, um estado do jogo, filtra as informações importantes, avalia possíveis caminhos ou soluções para alterar essa situação para uma que considera mais apropriada para sua vitória, decide por um desses caminhos e toma a ação que o leva a essa caminho. Em um jogo entre humanos, o adversário segue o mesmo ciclo [G. Xexéo, Eduardo Mangeli et al., 2021], como mostrado na Figura 5.2. A figura chama a atenção para o fato que, enquanto jogam, os jogadores se comunicam, socializando dentro ou fora da fantasia que o jogo cria.

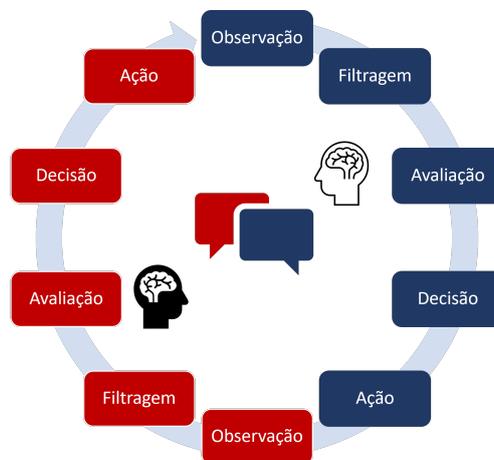


Figura 5.2. O ciclo do jogar para dois jogadores. Fonte: própria.

Se o adversário for um software, ou mesmo um artefato mecânico, o mesmo tipo de ciclo acontece, porém, podemos simplificar a ação do computador com um ciclo simples de detecção, processamento e *feedback*.

Nesse ciclo, o *designer* de jogos deve estar atento a oferta de decisões, e a dificuldade de realizar as ações desejadas. No Xadrez, basta mover uma peça, ou até mesmo falar a jogada para outra pessoa mover. Já em esportes, que são jogos, é necessária a habilidade física de realizar a ação decidida.

Devemos lembrar, também, que o jogo, na área de *Game Studies* é tratado como atividade voluntária [Huizinga, 1938–2001; Callois, 2001; G. Xexéo, Carmo et al., 2017]. Normalmente, o jogo educacional não é voluntário. Isso traz implicações em várias questões teóricas do jogo, como, por exemplo, a atitude lusória [Suits e Hurka, 2014], a atitude voluntária do jogador ao jogar e se submeter as regras do jogo.

A partir dessas análises, que podem ser aprofundadas e ampliadas, já é possível concluir que o estudo dos jogos de entretenimento não é suficiente para tratar de jogos educacionais, pois algumas premissas básicas do estudo de jogos de entretenimento são quebradas.

É necessário, então, construir uma base teórica específica para os jogos que têm outro objetivo que não apenas entreter, o que é feito em torno da denominação de **Jogo**

Sério, uma escolha no mínimo infeliz, mas que engloba todos os jogos criados com um propósito outro que não o de ser um jogo *per se*. Um nome geral mais adequado seria **Jogo com Propósito**. Os nomes específicos dados a sub-áreas dos jogos sérios, porém são mais esclarecedores: *newsgames*, jogos de persuasão, jogos de negócio, jogos empresariais, jogos educativos, jogos educacionais, *edutainment*, etc.

Um jogo sério, ou com propósito, é um artefato que usa a mídia artística dos jogos para passar uma mensagem, ensinar uma lição ou fornecer uma experiência. A palavra “sério” reflete o objetivo do jogo, não o conteúdo propriamente dito [Michael e Chen, 2006]. Abt [2002] diz que esses jogos “tem um objetivo explícito e cuidadosamente pensado”.

5.3.1. O Que São Jogos Educacionais

De forma sucinta, **um jogo educacional é um jogo criado com o propósito de apoiar o ensino-aprendizagem.**

Não devemos, porém, confundir o uso de jogos na educação com a gamificação da educação. A gamificação, uma adaptação do inglês *gamification*, é o uso de elementos de jogos para modelar, ou alterar, um processo ou comportamento. Um processo gamificado não necessariamente utiliza jogos.

Jogos educacionais têm muitas classificações e objetivos. Eles podem ser classificados pelo objetivo a que se destinam em:

- motivação para um processo de aprendizagem;
- transmissão de conhecimento;
- fixação de conhecimento;
- verificação de conhecimento;
- prática de habilidades por meio de simulações mais, ou menos, realistas, e
- criação de conhecimento por meio de processos de análise de acontecimentos.

Jogos educacionais podem ser criados para e usados em três contextos de aprendizagem diferentes:

1. aprendizagem formal, aquela que ocorre dentro de uma instituição de ensino ou no local de trabalho e é “ explicitamente concebida como uma atividade de aprendizagem” [CEDEFOP, 2014], intencional por parte do aprendiz.
2. aprendizagem não formal, aquela “integrada em atividades planejadas que não são explicitamente projetadas como atividades de aprendizagem” [CEDEFOP, 2014], mas intencional por parte do aprendiz, e
3. aprendizagem informal, aquela que ocorre de forma não intencional pelo aprendiz, no convívio diário, com a família e colegas, ou em atividades como brincadeiras e jogos [CEDEFOP, 2014; Kalmpourtzis, 2018].

Por exemplo, Sr. X [Oliveira, 2015], da Tamboro (Figura 5.3), jogo analisado por Taucei [2019], foi criado para ser usado tanto na aprendizagem formal quanto informal, podendo ser usado também na aprendizagem não-formal, apesar do tópico do jogo, equações do primeiro grau, não ser típico desse espaço. Já o Mapa do Tesouro, na Figura 5.4, também analisado por Taucei [2019], foi concebido para aprendizagem informal.



Figura 5.3. Tela do jogo Sr.X, da Tamboro, cujo objetivo é capturar o espião Sr. X, e a forma para isso é isolá-lo em um dos vagões, seguindo mecânicas que se equivalem as operações algébricas que são feitas para resolver uma equação de primeiro grau. Fonte: imagem obtida durante uma sessão do jogo Sr. X da Tamboro.

5.3.2. Distinção Entre Jogos Educacionais Exógenos e Endógenos

Com relação ao conteúdo e teorias que serão aprendidas pelos alunos ao jogar, os jogos educacionais podem ser exógenos ou endógenos.

Nos **jogos exógenos** os conflitos entre teoria, conteúdo e *game design* raramente surgem, uma vez que o conteúdo de aprendizagem e as mecânicas dos jogo estão separados, [Halverson, 2005]. Segundo Garris, Ahlers e Driskell [2002], em jogos exógenos é comum o conteúdo a ser aprendido ser inserido em uma estrutura e regras de um jogo pré-existente. Nesse caso o aprendizado pode ficar limitado memorização ou recordação de algum conhecimento. Um jogo de perguntas e respostas, conhecidos como jogos de trivía, pode ser usado com qualquer conjunto de conhecimentos, e, apesar de pouco sofisticados, é capaz de apresentar resultados positivos, principalmente em relação à motivação dos alunos.

Os **jogos endógenos**, de modo geral, focam em objetivos de aprendizagem mais complexos, que vão além da memorização, e fazem isso integrando o conteúdo de aprendizado a estrutura do jogo. Eles usam o que é conhecido como Retórica Procedural [Bogost, 2008], em que a mensagem é transmitida por meio das mecânicas do jogo. Esse tipo de jogos possuem, possivelmente, uma dose de simulação do mundo real, e implicam na existência de um modelo do mesmo, ainda que bastante abstrato.

O Mapa do Tesouro [Nascimento, G. B. Xexéo e Brito, 2019], cuja tela principal está na Figura 5.4, é um jogo educacional endógeno, construído usando técnicas de Retórica Procedural, criado por Pedro Nascimento dentro do LUDES³, já que visa ensinar a programação paralela, e ela está totalmente inserida nas regras e mecânicas do jogo.

Este tutorial trata especificamente da **criação de conceitos de jogos educacionais endógenos com o objetivo de tratar um problema no ensino-aprendizagem e com um público determinado**. Estes dois assuntos, o problema a ser tratado e a identificação do jogador/aluno, compõe o primeiro bloco do ENDO-GDC.

O leitor interessado em aprofundar os estudos em jogos educacionais e aprendi-

³<http://ludes.cos.ufrj.br>



Figura 5.4. Uma tela do Mapa do Tesouro, criado por Pedro Nascimento com a orientação e apoio dos membros do LUDES. Nela, piratas tentam encontrar o tesouro seguindo um programa feito pelo jogador. Na imagem, tela inicial com apenas um pirata. Fonte: própria.

zado baseado em jogos pode procurar textos como os de Plass, Meyer e Homer [2019], Prensky [2003], Abt [2002], e L. D. Costa [2008]. Muitos artigos do SBGames - Trilha Educação também apresentam excelentes estudos.

5.4. O Projeto de Jogos em Geral

Segundo Schell [2014], um jogo é composto de quatro elementos: mecânica, história, estética e tecnologia (Figura 5.5). A estética é como o jogo é percebido, suas formas, cores, cheiros. A tecnologia é como ele é realizado. A história, ou tema, algumas vezes opcional, é a narrativa associada ao jogo. Esses três elementos são encontrados não só em jogos, mas em outras mídias de entretenimento. As mecânicas são o grande diferencial. Elas são as ações que podem ser realizadas pelo jogador para influenciar o resultado do jogo.

No Xadrez, o tabuleiro e o formato das peças, e até mesmo o peso e o som que fazem ao tocar no tabuleiro, fazem parte da estética. A tecnologia pode ser apenas a madeira talhada e colada, mas também temos jogos de Xadrez eletrônicos, ou softwares. A mecânica é a alternância de jogadas entre os jogadores, o movimento de cada peça, o xeque e o mate. A história, porém, é quase nula, sem narrativa, havendo apenas um tema de uma batalha entre dois exércitos. Outros jogos, como os jogos solitários que encontramos em livros-jogos, podem ser basicamente compostos de narrativas, sendo a escolha da próxima página a única decisão.

Os quatro elementos são interdependentes, um servindo para melhorar e reforçar o outro, se o *design* for bem feito. Isso leva Schell a chamá-los da **Tétrade Elementar do Projeto de Jogos** (Figura 5.5).

Em outra abordagem, Hunicke, Leblanc e Zubek [2004] propuseram o modelo MDA, sigla que significa Mecânica, Dinâmica e Estética, apresentado na Figura 5.6. Para os autores, o projetista do jogo cria as **mecânicas**, os algoritmos e dados do jogo, e o jogador joga através de **dinâmicas**, a partir da experiência emocional que tem ao jogar, a **estética**. É importante frisar que os conceitos de mecânica e estética do MDA[Hunicke, Leblanc e Zubek, 2004] e da Tétrade Elementar Schell [2014] diferem, apesar de terem

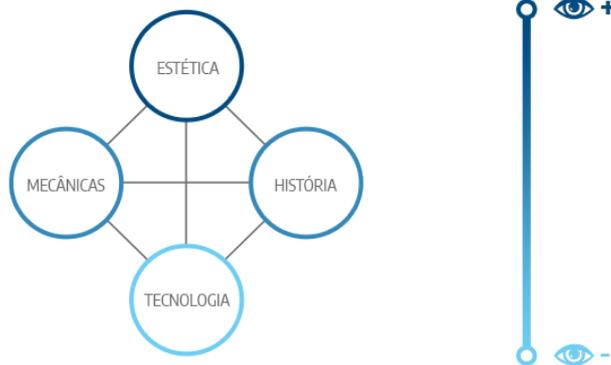


Figura 5.5. A Tétrade Elementar de Schell [2014], onde a estética é a parte mais visível. Fonte: inspirada em Schell [2014].

os mesmos nomes.

De acordo com os Hunicke, Leblanc e Zubek [2004], jogos são criados por desenvolvedores/designers e fruídos por jogadores. Ao trabalhar com jogos, é útil considerar essas duas perspectivas uma vez que ajuda a observar como pequenas mudanças em uma camada podem entrar em cascata nas outras. As diferentes perspectivas propiciam as seguintes perguntas:

- **do designer:** ao se implementar uma determinada mecânica, quais dinâmicas serão possíveis, e qual será a estética subjacente? Ambas estão de acordo com o que foi planejado pelo *Game Designer* ou equipe desenvolvedora?
- **do jogador:** quais dinâmicas devem estar presentes para que o jogador experimente uma determinada sensação? Que mecânicas (regras do jogo) devem ser desenvolvidas para tornar essas dinâmicas viáveis?

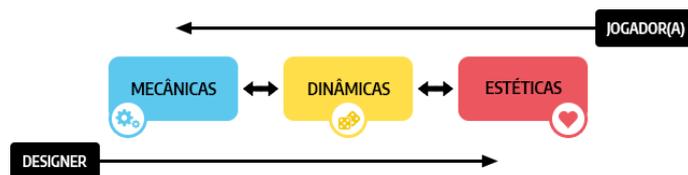


Figura 5.6. Modelo MDA (*Mechanics, Dynamics and Aesthetics*). Fonte: adaptado de Hunicke, Leblanc e Zubek [2004]

Pensar pela perspectiva do jogador incentiva um design orientado para a sua experiência ao invés do design orientado por recursos ou funcionalidades, [Hunicke, Leblanc e Zubek, 2004]. Dessa forma, as diferentes perspectivas apresentadas no MDA são importantes para ferramentas como o ENDO-GDC, pois podem auxiliar a equipe a confrontar as funcionalidades do jogo com a experiência que se deseja passar para o jogador logo no início do desenvolvimento do projeto. A estética no MDA é o equivalente à proposta de valor do ENDO-GDC de Osterwalder [2004], mas esse valor vem na forma de experiência e emoções para o jogador.

Apesar de ser um modelo simples e já ter sido usado por outro *canvas* — *Gamification Model Canvas* de Jiménez [2013] — o MDA não é um modelo adequado para jogos

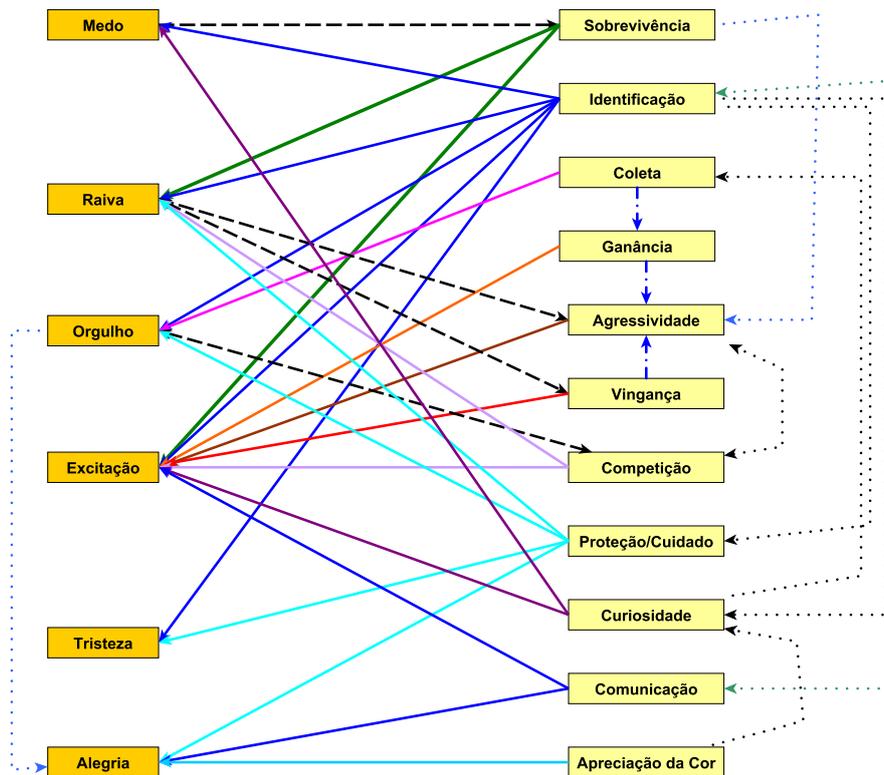


Figura 5.7. Interações entre emoções e instintos no *framework 6-11*. Fonte: adaptada de [Dillon, 2010]

sérios de acordo com Winn [2009]. Segundo o autor, jogos sérios são “jogos que vão além do *gameplay*”⁴ e seu desenvolvimento possui um conjunto de desafios únicos, como por exemplo, narrativas, experiência do usuário e componentes tecnológicos, os quais não são contemplados pelo modelo MDA. O autor propõe outro modelo, que veremos mais adiante.

Inspirado pelo MDA, Dillon [2010] propôs o *Framework 6-11*, de análise da experiência dos jogos, onde considerava que a Estética era melhor descrita por meio 6 emoções e 11 instintos, que se relacionam como mostrado na Figura 5.7.

As escolhas de Dillon são arbitrárias e podem causar polêmicas, mas são feitas sobre princípios razoáveis. As seis emoções — medo, raiva, orgulho, excitação, tristeza e alegria — são uma espécie de interseção entre vários autores que defendem a existência de emoções básicas [Dillon, 2010]. Já os instintos, indicados na Figura 5.7, que são acionados por eventos, já causam polêmica com o próprio nome, e o autor indica isso, já que há controvérsias se humanos tem instintos, pois podem controlá-los. Mas, novamente, seguindo princípios razoáveis, Dillon escolhe 11 instintos, ou características humanas acionadas por eventos, identificáveis e podem ser geradas por, ou gerar, as emoções básicas. Em ambas as escolhas, porém, fica claro que a intenção é usar o resultado no projeto e análise de jogos.

Além disso, Dillon [2010] criou um diagrama para a descrição de jogos, baseado

⁴*Gameplay* e mais um termo sem tradução na área de jogos.

nos relacionamentos entre os três elementos do MDA, onde os instintos ativam o uso das dinâmicas, que exigem mecânicas para serem realizadas. e os instintos também geram emoções.

A Figura 5.8 mostra um diagrama desse tipo, para o jogo Pac-Man. Nesse jogo, o pac-man é um círculo com uma boca que anda em um labirinto comendo pontos, pílulas, e frutas. No labirinto existem quatro fantasmas que perseguem e podem comer o pac-man, porém se ele come uma pílula, passa a poder comer os fantasmas, que passam a fugir dele, por um tempo específico. No diagrama, a estética, ou as sensações do jogador começam com a identificação e a vontade de colecionar pontos. Seguindo o caminho da identificação do jogador com o pac-man, ele terá medo dos fantasmas, o que implica no instinto de sobrevivência. Esse instinto sugere a dinâmica de escapar, que exige a mecânica de mover. Outros caminhos no diagrama mostram comportamentos distintos do jogador.

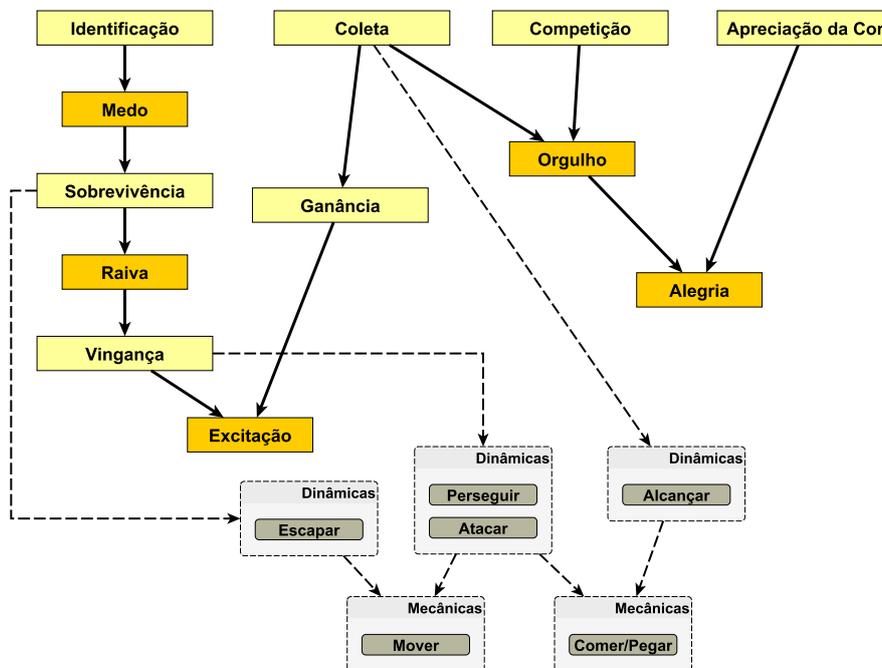


Figura 5.8. Um diagrama do *Framework 6-11* para o jogo Pac-Man. Fonte: adaptado de Dillon [2010]

5.4.1. Processos de Desenvolvimento de Jogos

A forma como jogos são desenvolvidos podem ser descritas formalmente como processos. Um processo “é uma coleção de atividades inter-relacionadas que são iniciadas em resposta a um evento e tem como objetivo alcançar um resultado específico para o consumidor do processo.[von Scheel et al., 2015]” Neste capítulo processos são descritos em BPMN [OMG, 2013].

Dzgoeva [2016] modela como processos cinco formas de desenvolvimento de jogos que são referências comuns na literatura, usando *procedure models* e Petri Nets [Schonthaler et al., 2012], incluindo os propostos por Salen e Zimmerman [2003] e Schell

[2014], seminais na área de design de jogos. Em todos os processos modelados há a exigência de existência prévia de um conceito de jogo, ou a criação do conceito do jogo. Do trabalho de [Salen e Zimmerman, 2003], Dzgoeva extrai um processo de desenvolvimento de jogos muito abstrato, apresentado na Figura 5.9, onde a existência do conceito do jogo é o evento que inicia o processo, sendo seguido um processo iterativo e incremental que pode ser classificado como de protótipos evolutivos [Pressman e Maxim, 2019]. Em outros processos, a criação de um conceito de jogo aparece no início do processo, podendo ser retomada em caso de um novo ciclo de melhoria do jogo.

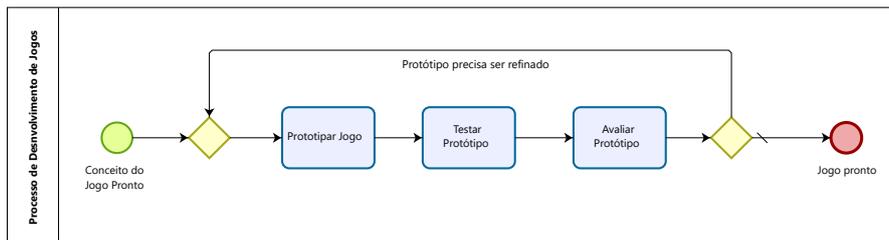


Figura 5.9. Processo de desenvolvimento de jogos segundo Salen e Zimmerman [2003]. Fonte: adaptado de Dzgoeva [2016] para o BPMN pelos autores

Basicamente todos os autores defendem que o desenvolvimento de jogos tem que ser iterativo e incremental, principalmente pela importância do *playtest*, que analisa a resposta do usuário ao jogo, como indicado por Winn [2009]. Jogadores sempre surpreendem o designer, pois buscam otimizar ao máximo as regras e mecânicas ao seu favor, possivelmente encontrando brechas e combos não imaginados *a priori*. Outras vezes, há dificuldade de compreensão de uma regra que pode parecer simples ao *game designer*, por seu conhecimento amplo dos jogos, mas é desconhecida do público específico.

O'Hagan, Coleman e O'Connor [2014] fazem uma revisão sistemática do processo de desenvolvimento de videogames, relatando que em 356 estudos primários estudados, onde foram identificados processos, 47% eram métodos ágeis e 53% eram híbridos, onde os dois diferem basicamente no peso das iterações no processo. Todos os métodos ágeis usam formas incrementais de desenvolvimento que passam pelos estágios de requisitos e design a cada iteração.

Marklund et al. [2019] a partir de 48 artigos, de uma coleção inicial de 2278, contendo dados sobre “desenvolvimento de jogos feito na indústria [de videogames], com alta qualidade de pesquisa e relevantes para a prática de desenvolvimento de jogos”, encontraram evidências que a “criatividade é alcançada através de um trabalho colaborativo, em um processo dirigido por testes, onde documentação e controle são atenuados”. Este trabalho também indica que a documentação mais citada é o *Game Design Document*, indicando também que técnicas como o compartilhamento de conhecimento e diálogos informais contínuos mantêm a visão criativa ao longo do desenvolvimento.

Existem muitas propostas de processos de desenvolvimento de jogos, tanto de entretenimento, quanto de sérios ou educacionais, descritas em diferentes níveis de abstração [O'Hagan, Coleman e O'Connor, 2014]. A maioria dos relatos analisados pelas nossas referências descreve o desenvolvimento de videogames. Não há consenso da melhor prática a ser seguida, apesar de algumas diretrizes gerais serem comuns, como a

necessidade de testar o jogo e de usar múltiplas iterações e trabalhar em incrementos [Winn, 2009; Marklund et al., 2019], e a viabilidade de usar métodos ágeis.

5.4.1.1. O Conceito do Jogo

A definição de “Conceito de Jogo” é bastante informal na literatura, basicamente podendo variar em uma linha até um documento razoavelmente estruturado, como um *Game Design Document* inicial.

Schell [2014] discute o desenvolvimento da ideia do jogo, que deve ser formalizada, definindo objetivos e restrições do jogo. Isso pode ser uma frase, como “um jogo de simulação de escalada para celulares”, que define um problema a ser resolvido, o que permite que a equipe busque e escolha soluções para essa ideia [Dzgoeva, 2016]. Já Bates [2004] fala de um conceito de alto nível, em várias sentenças que descrevem uma ideia, um documento de *pitch* seguido por um detalhamento das ideias iniciais [Dzgoeva, 2016]. Moore [2011] também cita documentos de *pitch* que são detalhados em *game proposals*, que descrevem o mundo e os elementos mais importantes do jogo.

A listagem poderia seguir, e vários exemplos podem ser dados, mas uma visão mais ampla seria dizer que o conceito do jogo é uma ideia geral do jogo, suficientemente detalhada de forma que o jogo possa ser criado a partir dela, e que seja possível decidir se o jogo criado está ou não alinhado ao conceito.

Desta forma, o ENDO-GDC será o próprio conceito do jogo.

5.5. Trazendo a Pedagogia para os Jogos

Nesta seção entramos cuidadosamente na área da Pedagogia, sendo os autores especialistas em Computação e no desenvolvimento de jogos. Nos apropriamos, então, de teorias amplamente aceitas, escolhidas por sua adequação ao tópico. É importante notar as escolhas feitas são limitadas, mas que representam uma seleção educada de práticas e teorias que são apenas um recorte de uma área muito mais ampla.

Uma definição possível de aprendizado, adequada ao contexto deste tutorial, é “o processo de criação de competências para resolver problemas” [Kalmpourtzis, 2018].

Competência é definida na Base Nacional Comum como Curricular (BNCC) “a mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas, cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho” [BRASIL, 2018]. Essa conceituação é uma construção que parte de três princípios originalmente propostos por Parry [1996] e amplamente aceitos na literatura, que no Brasil assumem a sigla CHA (Conhecimentos-Habilidades-Atitudes), somados a uma visão mais atual que incluem os valores.

Segundo Zabala e Arnau [2014], o “termo competência surgiu como consequência à necessidade de superar um ensino que, na maioria dos casos, reduziu-se a uma aprendizagem cujo método consiste em memorização”.

Competências são adquiridas pelo aprendizado em três contextos: o informal, o

não-formal e o formal [Kalmpourtzis, 2018]. O formal acontece nos ambientes dedicados a educação, como a escola. O não-formal e informal acontecem fora da escola, sendo que a não-formal acontece em ambientes privilegiados, como museus, e a informal acontece no dia a dia, com a família, etc. A educação formal e a informal também se contrapõem em algumas características, como a criação de currículos e ementas que acontece na formal, ficando a não-formal, em situação intermediária.

Analisando a competência por estes quatro aspectos, e os jogos por seu caráter motivacional e por características de simulação que devem ser usadas, podemos concluir que eles podem ajudar não só a aumentar o conhecimento, mas também a melhorar habilidade e atitude frente a um assunto, fornecendo experiências simuladas e ajudando a criar valores.

Assim, a seguir, é analisada uma forma de classificar as competências, a partir da Taxonomia Revisada de Bloom, e duas propostas de formas de compreensão e criação de jogos educacionais.

5.5.1. Taxonomia de Bloom

A taxonomia de Bloom é uma ferramenta pedagógica composta por um conjunto de três modelos hierárquicos, usados para classificar os objetivos de aprendizagem educacional em níveis de complexidade e especificidade. Os modelos abrangem os objetivos de aprendizagem nos domínios cognitivo (focado no conhecimento), afetivo (focado nas emoções e sentimentos) e psicomotor (focado nas habilidades motoras) Ferraz e Belhot [2010]. Para o desenvolvimento dos objetivos de aprendizagem do endo-GDC, utilizou-se o modelo cognitivo da Taxonomia Revisada.

Ambas as versões da taxonomia podem ser entendidas como uma pirâmide dividida em seis níveis, que indicam a complexidade cognitiva, do menos complexo, na base da pirâmide ao mais complexo, no seu topo, como pode-se observar na figura 5.10.

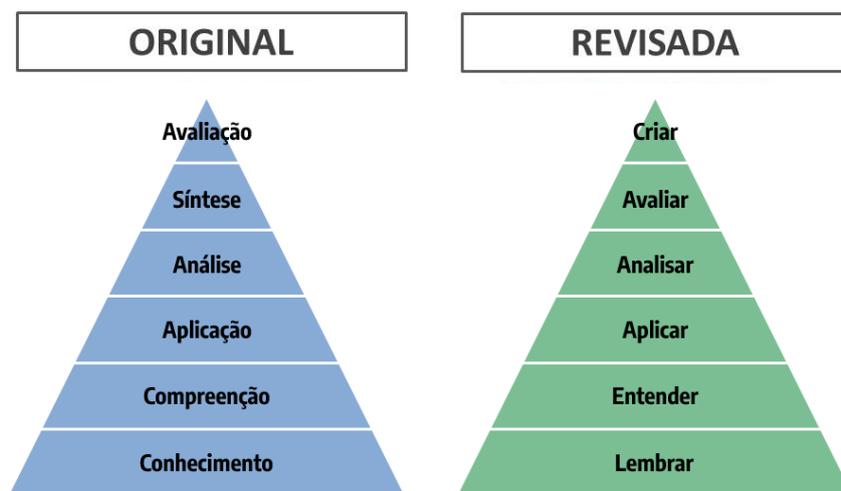


Figura 5.10. Taxonomia de Bloom Original e a Revisada

Algumas mudanças foram feitas na nova versão: três categorias foram renome-

adas, a ordem de duas foram trocadas e os nomes das categorias remanescentes foram alteradas para forma de verbos Krathwohl, [2002], para se ajustar a maneira como são usados nos objetivos.

Além das mudanças citadas acima, no modelo revisado o tipo de conhecimento a ser adquirido e o processo utilizado para a aquisição desse conhecimento foram separados em duas dimensões: de conhecimento e de processos cognitivos. O tipo de conhecimento passou a ser designado por substantivo e os processos cognitivos para atingi-los passaram a ser descritos por verbos, que compõem a pirâmide verde da figura 5.10 Krathwohl, [2002] e Trevisan e Amaral, [2016]. O objetivo desta mudança é facilitar a tarefa de definir com mais clareza os objetivos de aprendizagem.

5.5.2. DPE

O *framework* DPE, apresentado na Figura 5.11, foi criado como uma expansão do MDA para atender às necessidades do design de jogos sérios. A sigla é um acrônimo para *Design, Play and Experience*. O objetivo deste *framework*, é fornecer linguagem comum para discutir o design de jogos sérios, uma metodologia para analisar seu design e um processo para projetar um jogo sério para o aprendizado [Winn, [2009]].

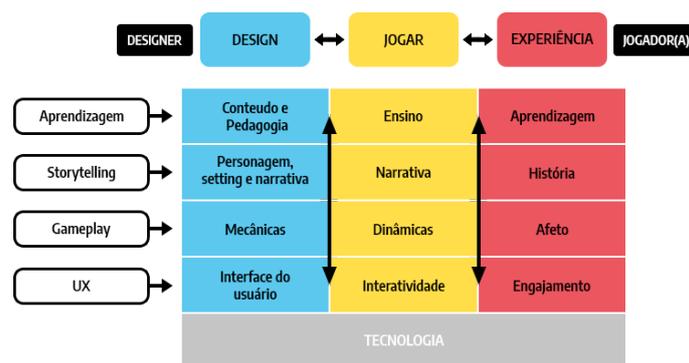


Figura 5.11. Modelo DPE (*Design, Play and Experience*). Fonte: adaptado de Winn [2009]

De maneira semelhante ao que ocorre no modelo MDA, o DPE também considera a relação existente entre o designer e o jogador. Segundo Winn [2009], o jogo é uma experiência mediada, onde a experiência que o jogador terá depende não somente do design do jogo como também do seu *background* cognitivo, social, cultural, e de experiências que ele teve no passado e também das que são geradas ao jogar. Assim, é essencial considerar o público-alvo do jogo ao longo do processo de design [Winn, [2009]].

O modelo é formado por quatro camadas: aprendizado, narrativa (*storytelling*), jogabilidade (*gameplay*) e experiência do usuário (UX - *user experience*) [Winn, [2009]].

Na camada de aprendizagem o *designer* projeta o conteúdo e a pedagogia, que poderá resultar no aprendizado desejado quando o aluno/usuário interagir com jogo. O Ato de jogar leva a um conjunto de resultados de aprendizagem, planejados ou não, derivados da experiência geral com o jogo. É nessa camada também que devem ser definidos os objetivos de aprendizado, que vão formar a base para a avaliação da eficácia da aprendizagem do jogo no jogador e servem de base também para o conteúdo e pedagogia [Winn, [2009]].

Na camada narrativa é desenvolvido a história do jogo. A narrativa, nos jogos, possui duas perspectivas: a história do designer e a história do jogador Rouse [2010]. A história do designer é a narrativa que é projetada no jogo (definir o cenário, fornecer engajamento e propósito do jogo e transmitir o conteúdo...). A história do jogador é experiência que ele teve com a narrativa do jogo, produto da história arquitetada pelo designer com as interações e escolhas realizadas pelo jogador durante o *gameplay*. Dessa forma, a história do jogador só é passível de ser analisada após um *playtest* [Winn, 2009].

Na camada de jogabilidade é definido o que o jogador pode ou não fazer, que escolhas que este pode fazer no mundo do jogo e as ramificações que estas escolhas terão no resto do jogo [Winn, 2009]. Ela é muito similar ao modelo MDA e é formada por três elementos: a mecânica, dinâmica e afeto. As mecânicas são as regras que vão definir o funcionamento do mundo do jogo, o que o jogador pode fazer, seus desafios e objetivos. As dinâmicas são os comportamentos resultantes quando as regras são instanciadas ao longo do tempo, durante o *gameplay*, com a influência das interações do jogador. As experiências resultantes, ou emoções derivadas no jogador, são os afetos (*affects*). Neste *framework* o termo *aesthetics* (estética do modelo MDA), comumente utilizados nas artes visuais para se referir a beleza física de algo é substituído pelo afeto (*affect*), um termo da psicologia que significa emoção ou desejo. O uso da palavra afeto procura dar mais ênfase a questão das emoções que se deseja emanar do jogador através do jogo.

A camada de experiência do usuário é a camada mais visível sob a perspectiva do jogador. O propósito da interface é tornar o entretenimento acessível e, no contexto de jogos sérios, é também criar um veículo capaz de realizar os resultados de aprendizado desejado. A interface engloba tudo o que o usuário vê, ouve e interage e como essa interação acontece, ou seja, o sistema de controle [Winn, 2009].

A última camada é a de tecnologia. Segundo os autores do *framework* algumas escolhas de design são mais dependentes da tecnologia do que outras. Enquanto algumas narrativas, mecânicas e elementos de aprendizado podem funcionar bem em jogos de tabuleiro ou cartas, em outros contextos isso não pode ser possível, caso o jogo sério necessite de uma simulação em tempo real, representação 3D, alguma narrativa mais complexa onde a história é mais sensível as decisões tomadas ao longo do *gameplay*, para citar alguns exemplos. Dessa forma, a tecnologia pode ser tanto um facilitador quanto um limitador do que será desenvolvido no jogo [Winn, 2009].

Cada camada apresentada do modelo tem influência sobre as outras camadas – indicado pelas setas pretas verticais, na figura 5.11. De acordo com Winn [2009]:

para a maioria dos jogos sérios, a aprendizagem é o aspecto mais importante e, geralmente, o menos maleável. A narrativa muitas vezes está amarrada ao conteúdo de aprendizagem e, portanto, deve ser abordado em seguida, mas é, geralmente, muito mais maleável. As camadas de jogabilidade e experiência do usuário são as mais maleáveis e muitas vezes devem se adaptar ao aprendizado e narrativa.... As decisões em níveis inferiores e descobertas em *playtesting* influenciarão o projeto em todas as camadas...

O modelo DPE, fornece uma estrutura para decompor os elementos de design do jogo sério, entretanto alguns desses elemento só poderão ser descritos ou analisados

após um *playtest*. Esses elementos associados ao *playtest* não entram no ENDO-GDC, mas são informações importantes de serem armazenadas em um contexto de processo de desenvolvimento do jogo, já que podem influenciar o preenchimento do ENDO-GDC em uma nova interação de equipe.

5.5.3. Jogos e Aprendizado: *Input Process Outcome*

Em seu artigo “*Games Motivation and Learning*”, Garris, Ahlers e Driskell [2002], apresentaram o modelo *input-process-outcome* (figura 5.12) para jogos instrucionais e de aprendizagem que é baseado no modelo clássico de aprendizagem.

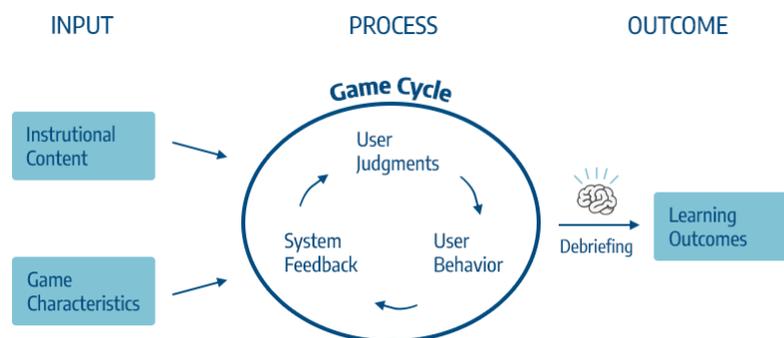


Figura 5.12. *Input-process-outcome*. Fonte: adaptado de Garris, Ahlers e Driskell [2002]

Segundo os autores, o modelo tradicional de aprendizado *input-process-output* enfatiza o aprendizado em uma única tentativa, onde o aluno que executa a tarefa uma única vez. No modelo *input-process-outcome*, determinadas características (*game characteristics*) ou recursos do jogo (fantasia, regras/objetivos, estímulo sensorial, desafio, mistério e controle) acionam o ciclo de jogo onde essas características podem levar a certos julgamentos (*user judgments*) do jogador sobre o *game*, se este é divertido, interessante e/ou envolvente; esses julgamentos levam a determinados comportamentos (*user behavior*) como maior persistência ou tempo na tarefa; esses comportamentos resultam de um *feedback* do sistema sobre o desempenho do jogador no contexto do jogo. Na medida que o emparelhamento do conteúdo instrucional (*instructional content*) com os recursos apropriados do jogo é bem sucedido, o ciclo resulta em um recorrente jogar automotivado. Esse engajamento ao jogar leva a conquista dos objetivos de treinamento e resultados de aprendizado específicos. O link entre o ciclo de jogo e os resultados de aprendizado é feito pelo processo de *debriefing*.

O *debriefing* é a revisão e análise dos eventos que aconteceram no jogo, ou seja, é a ligação entre o que é representado na simulação/experiência do jogo e o mundo real. Ele permite que os participantes façam paralelos entre eventos de jogos e eventos do mundo real. Em jogos "puros", esse link não precisa ser feito, como o jogo existiria dentro de seus próprios limites e ser jogado para seu próprio bem [Garris, Ahlers e Driskell, 2002]. Os *feedbacks* providos pelo jogo têm um papel fundamental na construção desse *link*.

De acordo com a visão dos autores a interação com o game faz com que jogador construa ativamente o conhecimento através da experiência com o jogo. O estágio

final que se procura alcançar, ao utilizar jogos para treinamento ou ensino, é ter um aluno motivado. De acordo com Skinner e J. Belmont [1993], esses tipos de alunos são entusiasmados, interessados, focados e engajados, gostam do que estão fazendo, tentam repetidas vezes, persistem ao longo do tempo e possuem motivação intrínseca e embora sejam fáceis de reconhecer, são difíceis de encontrar e de criar.

5.5.4. Os Modelos Teóricos e a Organização do ENDO-GDC

Os modelos teóricos apresentados, além de auxiliarem na escolha dos elementos de jogos que estão presentes no Endo-GDC, ajudaram organizar e agrupar suas seções e trazem aprendizados importantes para a concepção de jogos.

O MDA mostra os elementos centrais que constituem o *gameplay*, as mecânicas, dinâmicas e estéticas, e como eles se influenciam. Mostra também a importância de se observar o jogo a partir da perspectiva do jogador, incentivando um design orientado para a sua experiência.

O DPE, por sua vez, parte dos elementos tidos como centrais aos jogos — mecânicas, dinâmicas e estéticas — que no modelo é chamada de afeto — e é um modelo focado em jogos sérios. O modelo procura descrever como esses vários elementos estão relacionados, tanto no nível da camada a que pertencem, quanto entre as camadas. Além disso, ressalta a importância da interface do usuário como um elemento que ajuda atingir os resultados de aprendizado planejados para os jogos Winn, [2009]. As camadas apresentadas ajudaram a agrupar as seções do ENDO-GDC em blocos e temas específicos. O modelo ainda consegue preservar a perspectiva do usuário de do jogador, como ocorre no MDA.

A téttrade elementar apresenta quatro elementos que constituem um jogo: mecânicas, estética, história e tecnologia. O autor do modelo enfatiza a importância de se equilibrar esses elementos e como a alteração em um elemento pode afetar todos os outros.

Já o modelo *input-process-outcome* foca mais no processo de aprendizado através do jogo do que nos elementos que o formam. Apesar disso, há aspectos importantes nesse modelo que podem ser considerados para jogos educativos, como o conteúdo pedagógico, sistema de *feedback* e resultados de aprendizagem.

Há uma interseção dos elementos desses modelos (tabela 5.1). Mesmo não possuindo definições exatamente iguais, essa interseção mostra a relevância desses elementos e por isso foram usados na composição das seções do Endo-GDC. Na sua segunda versão o Endo-GDC teve a seção tecnologia removida e substituída pela seção restrições.

Além de apresentarem vários elementos, os modelos explicitam relações e influências que esses elementos possuem entre si e foram usadas como uma das referências para organizar as seções e agrupa-las em blocos. As características e base teórica para organizar e agrupar as seções de um canvas para jogos pode ser vista em Taucei, L. Costa e G. Xexéo, [2018] e Taucei, [2019]

Intercessões entre modelos			
Modelos	DPE + MDA + Tétrade Elementar	DPE + Tétrade Elementar	Input-processes-outcome + DPE
Elementos da interseção	Mecânicas e Estética	Mecânicas, Estéticas (afeto), História e Tecnologia	Conteúdo pedagógico e Resultado de Aprendizado

Tabela 5.1. Interseção entre os elementos presentes em cada modelo teóricos sobre jogos

5.6. O ENDO-GDC

Esta seção é a parte principal do tutorial. Ela apresenta o ENDO-GDC, explicando o que ele é e detalhando um passo a passo para sua utilização.

O objetivo do ENDO-GDC é servir como ponto focal de um grupo de trabalho que busca como objetivo, em uma ou poucas sessões, definir o conceito de um jogo educacional endógeno. Além disso, ele será o registro das decisões desse grupo, servindo também como objeto de discussão ao longo do desenvolvimento do jogo. Na prática, **o ENDO-GDC é o conceito do jogo.**

Como descrito anteriormente na sub-seção [5.3.2](#), o termo endógeno identifica que o conteúdo educacional e as regras interagem e são coesos, ao contrário de jogos exógenos onde as regras e o conteúdo são independentes. Isso só é objetivo com o esforço dedicado da equipe, porém o ENDO-GDC fornece constantemente o quadro que permite a execução e avaliação desse objetivo.

O **conceito do jogo**, como vimos na sub-seção [5.4.1](#), é parte essencial dos processos de desenvolvimento de jogos. Porém, uma documentação longa e estática não seria adotada pela academia ou indústria, como mostra a prática relatada por O'Hagan, Coleman e O'Connor [\[2014\]](#) e Marklund et al. [\[2019\]](#). Além disso, é necessário manter a visão do jogo ao longo do projeto de forma transparente para a equipe [\[Marklund et al., 2019\]](#), porém sujeita a modificações advindas do *play-testing* [\[Winn, 2009\]](#).

Analisando a cultura de desenvolvimento de jogos nos últimos 10 anos, se algum método é usado, o que em muitos casos parece não ser a realidade [\[Marklund et al., 2019\]](#), há grandes chances que esse método seja ágil ou híbrido [\[O'Hagan, Coleman e O'Connor, 2014\]](#); [Marklund et al., 2019\]](#).

Os canvas têm sido usados, principalmente em métodos ágeis, para garantir o acordo entre as partes, iniciar o engajamento, manter um registro transparente e facilmente atualizável de processos e produto, e facilitar a comunicação [\[Osterwalder e Pigneur, 2010\]](#); [Taucei, L. Costa e G. Xexéo, 2018\]](#). Eles fazem parte e disseminam a cultura ágil e enxuta, que se espalha por todas as indústrias.

Dessa forma, o uso de um canvas para facilitar a criação, manutenção e comunicação do conceito de jogos encontra apoio na indústria [\[Taucei, L. Costa e G. Xexéo, 2018\]](#), e utilizá-los para o caso educacional endógeno é uma alternativa atraente e viável, como

demonstrado por Taucei [2019] na proposta do ENDO-GDC.

5.6.1. O que é o ENDO-GDC

O ENDO-GDC, apresentado na Figura 5.13, é um canvas que, por sua natureza, permite um processo para o desenvolvimento do conceito do jogo. Nesse processo, são levantadas informações que criam um quadro ao redor dos principais conceitos: as mecânicas, as dinâmicas e a estética. Dessa forma, ao imaginar o jogo, a equipe responsável pela criação do conceito é, ao mesmo tempo, limitada e habilitada para usar esse quadro como alavanca para pensar e definir como o jogo deve intervir na situação descrita.

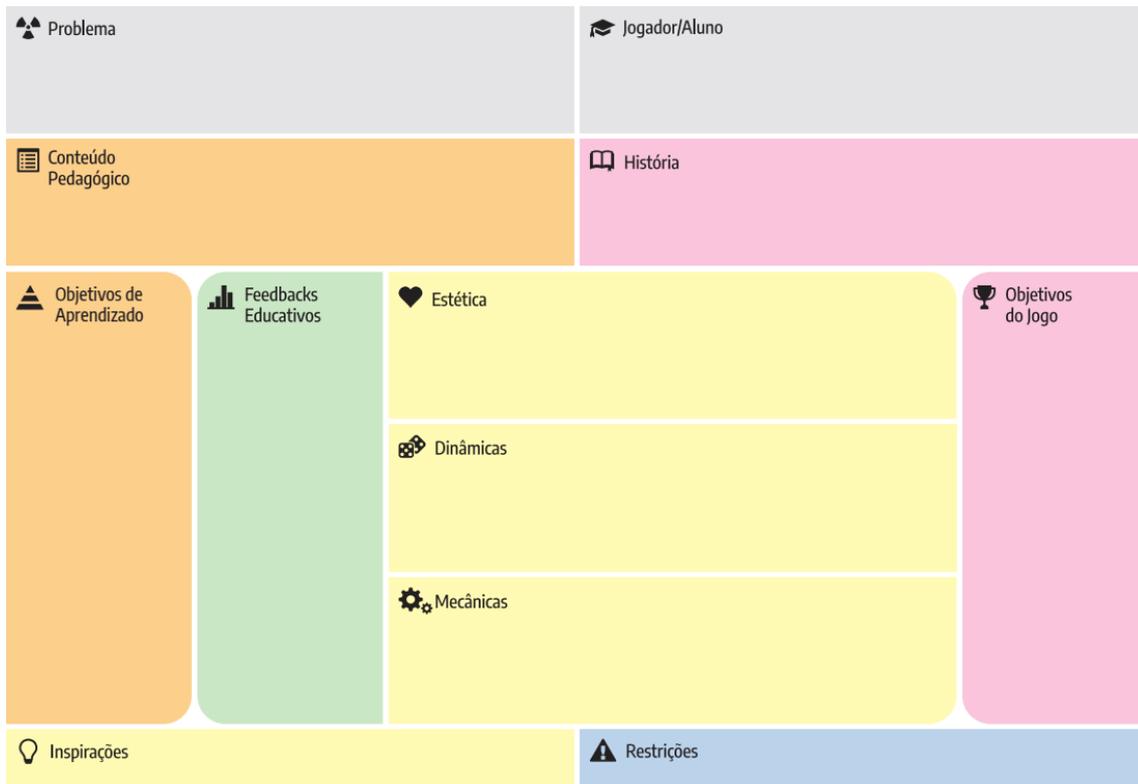


Figura 5.13. O canvas de projeto de jogos educacionais ENDO-GDC

O ENDO-GDC é formado por seis blocos, indicado por cores, que estão ligados às etapas que compõe o projeto de um jogo sério, como apresentado por Winn [2009]:

- **contexto:** com as seções problema e jogador/aluno, em cinza;
- **aprendizado:** com as seções conteúdo pedagógico e objetivos de aprendizado, em laranja;
- **narrativa:** com as seções história do jogo e objetivos do jogo, em rosa;
- **gameplay:** com as seções mecânicas, dinâmicas, estética, e inspirações em amarelo;
- **experiência do usuário:** com a seção *feedbacks* educativos em verde, e
- **restrições,** formada pela seção restrição em azul.

Além do canvas propriamente dito, o ENDO-GDC também usa cartas que ajudam a guiar a equipe. Essas cartas fornecem de maneira rápida a explicação da teoria e a

possibilidade de serem usadas em dinâmicas criativas. Essas cartas serão descritas no momento do seu uso.

Outra característica construída no ENDO-GDC é sua organização em duas partes: na esquerda são tratados assuntos relacionados ao aprendizado, e na direita assuntos relacionados ao jogo e a motivação. O *gameplay*, intencionalmente, cruza as duas partes.

5.6.2. Casos onde o ENDO-GDC pode ser aplicado

O ENDO-GDC foi criado para que o jogo criado possa superar um problema detectado em um processo de ensino-aprendizagem. Esse problema, por exemplo, pode ser a falta de motivação para um tema, a dificuldade de compreender algum conceito abstrato, ou o “medo” de uma matéria.

Por que precisamos de um problema? Na prática, porque não há motivo para desenvolver um jogo educacional se não há uma necessidade de melhorar a prática de ensino-aprendizado. Isso seria uma perda de tempo. Além disso, queremos deixar bem claro os benefícios que o jogo vai trazer no contexto específico em que será jogado.

O jogo pode ser criado em várias condições. Um professor, sozinho, pode usar o ENDO-GDC como guia para um processo de criação de um jogo simples, ou pode fazê-lo com outros professores ou um especialista em jogos. Empresas podem usá-lo para criar jogos que serão vendidos ao mercado, ou incluídos em uma plataforma educacional. Nenhum projeto é grande ou pequeno para o ENDO-GDC, porém ele é feito para criar um jogo, não oferecendo capacidades para criar uma sequência de jogos, a não ser que sejam na prática fases de um mesmo jogo.

Um uso interessante do ENDO-GDC pode ser fazê-lo com os próprios alunos que são o objetivo do projeto, para que criem um jogo que resolva uma dificuldade encontrada para o aprendizado. Essa dinâmica, porém, exige uma preparação anterior do professor, e implica no uso de técnicas de Aprendizado Baseado em Projetos [Larmer, Mergendoller e Boss, 2015]. Também, nesse caso, é possível que não seja adequado exigir que o jogo seja endógeno, devida a dificuldade adicional de criar um jogo desse tipo. O foco deixa então de ser o jogo e passa a ser o processo de criação do jogo, que inclui a elicitación do conhecimento necessário.

5.6.3. O Processo do ENDO-GDC

O ENDO-GDC deve ser usado dentro de um processo iterativo e incremental de desenvolvimento de jogos. Inspirados na descrição de alto nível que Dzgoeva [2016] faz da descrição de Salen e Zimmerman [2003], descrita na Figura 5.9, supomos que o ENDO-GDC é usado no processo descrito na Figura 5.14, em BPMN⁵, e é bastante genérico. Evitamos o uso da palavra protótipo, que já determina algumas técnicas específicas, e nos alinhamos também com a ideia de Winn [2009] sobre a necessidade de *playtesting* e avaliação, o que leva ao desenvolvimento iterativo e ao questionamento constante das decisões, inclusive ao conceito do jogo.

Para realizar o ENDO-GDC sugerimos um processo, baseado nos passos de Winn

⁵BPMN considera que um processo é uma sequência de atividades, que podem ser tarefas atômicas, ou sub-processos [OMG, 2013]

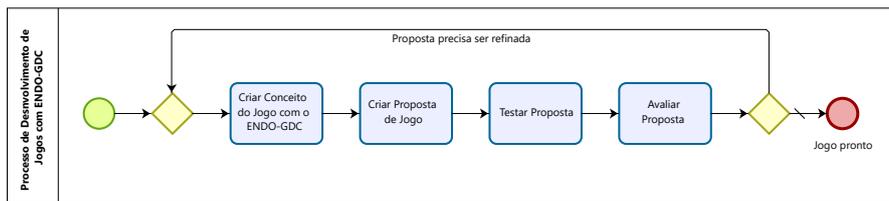


Figura 5.14. Inserção do ENDO-GDC no processo geral de criação de um jogo.
Fonte: imagem do autor.

[2009]. Esse processo não é, porém, uma camisa de força. Ele indica uma ordem preferencial, e também se apresenta deve ser usado na forma uma sequência de iterações que vão acontecer até que os designers estejam satisfeitos com o jogo proposto.

A Figura 5.15 dá uma visão geral do processo. Para criar o ENDO-GDC então são necessários pelo menos 7 atividades, sendo quatro subprocessos e 3 tarefas. Ao final de cada execução completa do ENDO-GDC pode ser necessário voltar ao início, por se constatar que não foi atingido um conceito adequado de jogo.

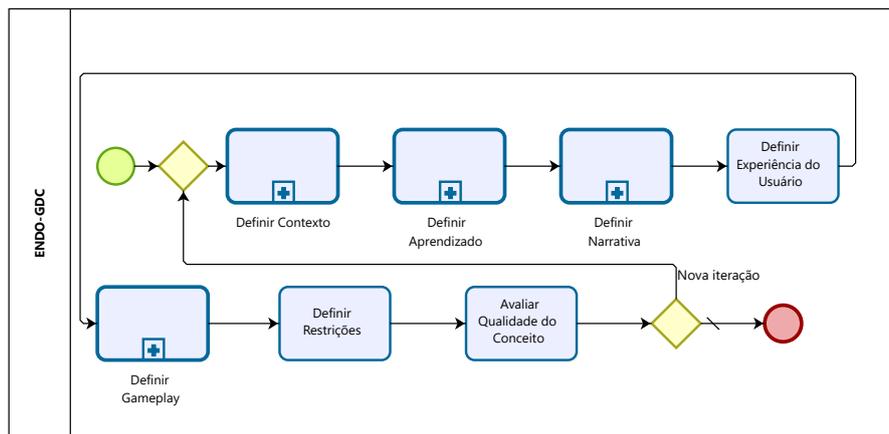


Figura 5.15. O processo de criação do ENDO-GDC descrito em BPMN. Fonte: Imagem do autor.

A primeira atividade é o subprocesso de Definir Contexto. Isso é feito em três passos, como descrito na Figura 5.16. É preciso definir o problema, ou seja, por que se fará o jogo, e então definir quem serão os alunos/jogadores.

Nesta parte, que é na prática uma justificativa para o projeto de criação do jogo, é importante um fechamento explícito, que garanta a consistência e completude do contexto para a iteração ocorrendo no momento. Começar um projeto qualquer sem uma definição precisa do porquê do projeto existir pode levar a caminhos inadequados, onde até pode se alcançar um produto que parece ser de qualidade, porém não serve para o propósito para o qual foi criado.

Após a definição do contexto, é necessário definir o aprendizado. Isso é feito em dois momentos: a definição do conteúdo pedagógico, e a definição dos objetivos de aprendizado, que indicam, segundo a Taxonomia Revisada de Bloom [Krathwohl, 2002], o nível de competência a ser alcançado em tópicos específicos do conteúdo pedagógico.

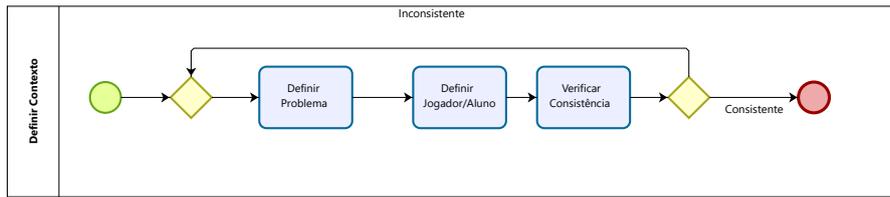


Figura 5.16. Definir contexto, um subprocesso do ENDO-GDC. Fonte: imagem do autor.

Este subprocesso está descrito na Figura [5.17](#).

O aprendizado só pode existir se o problema foi definido, já que é o problema que será usado para limitar tanto o escopo do conteúdo, como para a escolha dos objetivos de aprendizado.

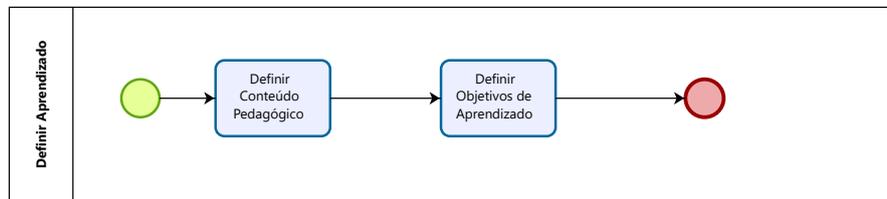


Figura 5.17. Definir aprendizado, um subprocesso do ENDO-GDC. Fonte: imagem do autor.

Com o contexto e os requisitos de aprendizado definidos, é possível criar uma narrativa, composta de uma história propriamente dita, onde ocorreram eventos que são objetivos e subobjetivos do jogo.

Novamente, desenvolver a narrativa exige as duas atividades anteriores. O contexto vai ajudar a encontrar temas de interesse ao jogador, e o conteúdo pode inclusive forçar o posicionamento da história no tempo e espaço.

Alguns jogos podem ser abstratos e não ter uma história, mas isso pode resultar em um engajamento mais baixo. Sugerimos que haja pelo menos um tema, uma inspiração para gráficos e sons. A Figura [5.18](#) ilustra esse subprocesso.

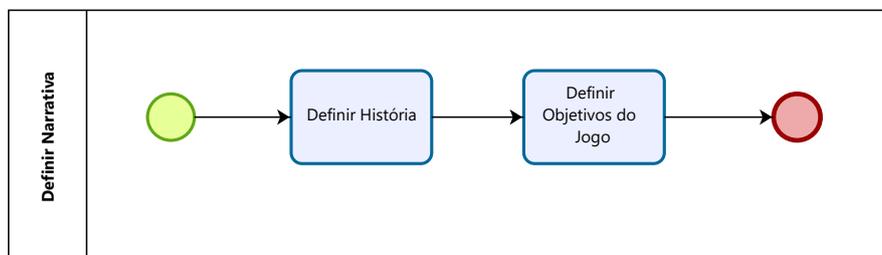


Figura 5.18. Definir a narrativa é um trabalho reconhecido do game designer. Fonte: imagem do autor.

À definição da história se segue a definição da experiência do usuário/jogador, nesse caso com o objetivo de definir os feedbacks educacionais que ele deve receber.

Nesse ponto, grande parte do ENDO-GDC está definido, fornecendo um quadro para a definição do *gameplay*.

Nesse momento a discussão é mais livre em relação a ordem, e as quatro atividades possíveis são executadas em paralelo, como mostra a Figura 5.19.

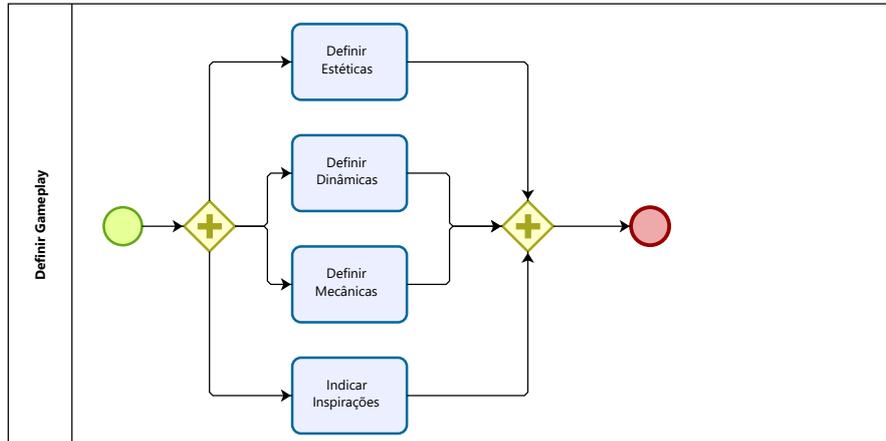


Figura 5.19. O subprocesso Definir Gameplay para a criação do ENDO-GDC.
 Fonte: imagem do autor.

Finalmente, deve ser avaliado o conceito completo e verificado se há necessidade de refiná-lo de alguma forma, para aumentar a consistência ou jogabilidade, por exemplo.

5.7. As Etapas do ENDO-GDC

As próximas sub-seções descrevem as etapas do ENDO-GDC.

5.7.1. Problema

Como dito na sub-seção 5.6.2, o ENDO-GDC é indicado para casos onde há um, ou mais, problemas de ensino-aprendizado. Esses problemas devem ser então identificados e relatados nessa seção.

Nela podem ser incluídos, mas não estão limitados a, itens como:

- dificuldades específicas das áreas do conhecimento, como o pensamento abstrato necessário a Matemática;
- falta de motivação dos alunos para temas e áreas específicas, como História do Brasil;
- problemas de formação anterior que exigem uma intervenção para possibilitar a ementa prevista;
- problemas de comportamento, falta de atenção, falta de companheirismo da turma, e ainda muitos outros.

Dependendo da realidade que causou o projeto o problema descrito vai ser um ou outro. Eles devem ser específicos ao escopo sendo analisado. Exemplos específicos de problema são:

- uma turma foi formada com alunos repetentes que já sabem parte da matéria, e

precisam completar esse conhecimento;

- um tema específico de Matemática apresenta dificuldade para os alunos em geral⁶, ou
- alunos precisam entender melhor o funcionamento de um mecanismo, como os juros compostos, e isso poderia ser feito por meio de um jogo digital que permitisse manipular o mecanismo.

É importante estar atento se um problema levantado não é apenas um sintoma de outro problema. Tentar tratar um sintoma pode trazer resultados de baixo desempenho. Se houver tempo disponível, é possível usar técnicas de análise de problemas, com o Diagrama Espinha de Peixe[Barbosa, s.d.], ou os Cinco Porquês[Gray, S. Brown e Macanuf, 2010].

Ao ser identificada uma causa de um problema, não é necessário retirar o problema original do ENDO-GDC, já que sua presença ajuda a entender o contexto do projeto, porém o foco principal considerar a causa.

5.7.2. Jogador/Aluno

Os Jogadores são o público alvo e o motivo do jogo existir. De acordo com Winn [2009] experiência que o jogador terá depende não somente do design do jogo como também do seu *background* cognitivo, social, cultural, e de experiências que ele teve no passado.

Assim, jogadores diferentes podem ter experiências totalmente distintas para uma mesmo jogo. Dessa forma, entender o jogador alvo do game é importante para aumentar as chances de desenvolver um jogo divertido e motivante, e seu perfil deve ser considerado durante todo o processo de design do jogo, [Winn, 2009]. A suposição que é possível criar um jogo que atende a todos está errada. Além disso, tanto o conteúdo pedagógico quanto os objetivos de aprendizado devem ser coerentes com o público-alvo.

Na seção do Jogador/Aluno é descrito quem é o público alvo do jogo, ou seja conjunto de pessoas que irão jogar e aprender através *gameplay*. Não estamos aqui discutindo pessoas, mas sim arquétipos e grupos típicos, ou proto-personagens.

Alguns dados são importantes serem definidos:

- Faixa etária: interesses, cognição e capacidade motora mudam com a idade [Schell, 2014].
- Profissão: qual a profissão desses jogadores alunos, caso sejam mais velhos.
- Nível escolar: o nível escolar ajuda a saber qual conteúdo pedagógico pode ser melhor explorado no jogo. Uma referência nacional para isso é Base Nacional Comum Curricular.
- Cenário de uso: qual o local que os usuários irão interagir com o jogo? Sala de aula? Casa? Em trânsito? Esses locais possuem algum tipo de restrição .

Cada uma dessas informações citadas acima pode ser escritas em um post-it separado, se o público alvo for único, ou em fichas por jogador, se o público alvo for maior.

Uma técnica interessante é buscar descrever personagens que são estereótipos radicais, a fim de colocar em questão a adequação destes jogadores ao resto da concepção

⁶Esse foi o caso do jogo para celulares Sr.X, que foi analisado como exemplo pela dissertação de Taucei

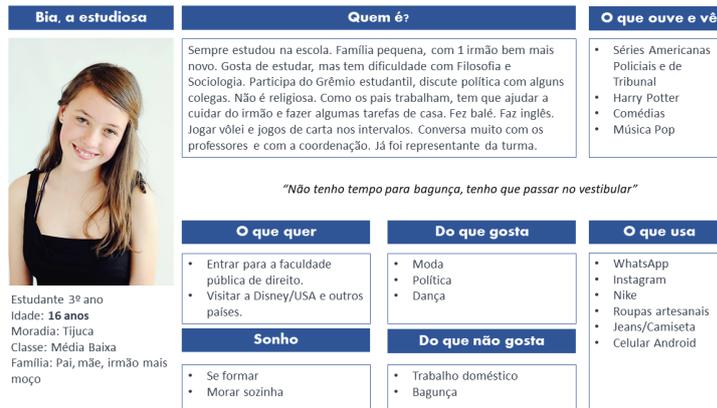


Figura 5.20. Exemplo de um *user persona*. Fonte: própria.

do jogo.

Ferramentas adicionais que podem ser usadas nesse caso, ou a partir da criação do ENDO-GDC, são o mapa de empatia[Gray, S. Brown e Macanufe, 2010] ou um *user persona*[D. M. Brown, 2010], exemplificado na Figura 5.20. Ambas tentam descrever um proto-usuário de um sistema ou produto, no caso, o jogo.

Ao terminar a etapa do jogador, a equipe pode perceber novos problemas ou desafios a serem tratados. Não há problema em, agora, analisar a concordância entre as duas etapas, e corrigir discrepâncias encontradas. Tanto pode ocorrer que nenhum jogador apresente um problema imaginado, ou que um jogador não apresente nenhum problema, ou que o jogador apresente um problema que ainda não foi listado.

5.7.3. Conteúdo Pedagógico

Nesta etapa devem ser levantada uma listagem de assuntos e conceitos que serão abordados pelo jogo e devem ser coerentes com o público alvo, definidos na seção jogador/aluno. Assim devem se listados:

- Área do conhecimento.
- Tópicos da área do conhecimento que serão abordados no jogo.
- Conteúdo que se deseja ensinar e está relacionada a área do conhecimento.

Cada uma das informações acima devem vir em um post-it separado na seção.

A descrição do conteúdo pedagógico é importante, pois os objetivos de aprendizado serão elaborados com base neste conteúdo. Além disso, é importante que os assuntos que serão lecionados tenham ligação como os diferentes elementos que formam o jogo, uma vez que pode reforçar o aprendizado e permitir explorar o ciclo de jogo, presente no modelo de Garris, Ahlers e Driskell [2002]

Nessa etapa, é importante manter o contexto criado pelas etapas anteriores. O conteúdo deve atender ao jogador, e deve apresentar o desafio definido no problema.

Aqui, se forem detectados outros problemas, deve se decidir se o conteúdo será mantido ou as etapas anteriores serão alteradas.

5.7.4. Objetivos de Aprendizado

A definição dos objetivos de aprendizagem é importante para saber se o jogo educacional atingiu o seu propósito, que deve ir além do puro entretenimento do jogador, de acordo com Abt [2002], Susi, Johannesson e Backlund [2007] e Michael e Chen [2006].

Nesta seção é descrito o que se espera que o jogador/aluno aprenda ao jogar. Existem três categorias amplas de resultados de aprendizagem segundo Garris, Ahlers e Driskell [2002] e Ferraz e Belhot [2010]: cognitivo, afetivo e psicomotor (baseado em habilidade). Nesta seção deve ser detalhado o tipo de resultado de aprendizagem desejado e o conteúdo pedagógico ao qual objetivo de aprendizagem está relacionado.

Além de ajudar a saber se o jogo projetado atingiu o seu propósito, os objetivos de aprendizagem servirão com um guia para construção do jogo, uma vez que as mecânicas, dinâmicas, estéticas e o *feedback* que devem ser desenvolvidas pensando-se em como alcançar esses objetivos.

Como será visto nas seções adiante, esses quatro elementos são uma simplificação do ciclo de jogo proposto por Garris, Ahlers e Driskell [2002], visto que não englobam os julgamentos e comportamentos do usuário, que só podem ser analisados em uma sessão de *playtest*.

O Endo-GDC focou na elaboração dos objetivos de aprendizagem relacionados a cognição, utilizando a Taxonomias Revisada de Bloom como sugerido por [Winn, 2009].

De certa forma, os objetivos de aprendizagem limitam e esclarecem o que está sendo esperado dentro do conteúdo pedagógico. Por exemplo, o conteúdo pedagógico pode ser Declaração da Independência, mas o objetivo, de acordo com a Taxonomia de Bloom, pode ser lembrar dos personagens e entender os motivos.

Para auxiliar os usuários do Endo-GDC, na elaboração dos objetivos de aprendizagem, são usadas cartas que descrevem a hierarquia de Bloom.

5.7.4.1. Cartas dos Objetivos de aprendizagem

Para auxiliar o preenchimento dos objetivos de aprendizagem, foram desenvolvidas cartas baseadas na Taxonomia Revisada de Bloom, Krathwohl [2002]. As cartas estão divididas nas de acordo com as dimensões da versão revisada da taxonomia: de conhecimento e de processos cognitivos.

A dimensão de conhecimento gerou quatro cartas baseadas em Krathwohl [2002] e Ferraz e Belhot [2010], organizadas de acordo com o nível de abstração do conhecimento (figura 5.21), do mais concreto 'A' ao mais abstrato 'D':

- **Factual:** os elementos básicos que um estudante deve saber para se familiarizar com uma disciplina ou resolver problemas nela. Fatos que não precisam ser entendidos ou combinados, apenas reproduzidos como apresentados.
- **Conceitual:** conhecimento das inter-relações entre os elementos básicos dentro de uma estrutura maior que permitem que eles funcionem juntos.
- **Procedural:** conhecimento de "como fazer algo", métodos de investigação e critérios para usar habilidades, algoritmos, técnicas e métodos



Figura 5.21. Cartas de Bloom da dimensão do conhecimento e carta de auxílio para “montar” o objetivo de aprendizado. Fonte: própria (2019)

- **Metacognitivo:** relacionado ao reconhecimento da cognição em geral e da consciência da amplitude e profundidade de conhecimento adquirido de um determinado conteúdo.

Em cada uma das cartas da dimensão do conhecimento (figura 5.21), temos a definição do tipo de conhecimento (definições acima); tipo de conhecimento do qual se refere (na parte mais inferior da carta), formato da sentença a ser formada (meio da carta) para construção do objetivo de aprendizado. A sentença, para formar o objetivo de aprendizado, tem a seguinte estrutura: verbo no infinitivo, que define a categoria cognitiva (carta de processo cognitivo); tipo de conteúdo, relacionado ao conteúdo pedagógico (objeto - o que será ensinado); processo cognitivo, que é o verbo no gerúndio, listados na seção de gerúndios parte inferior das cartas de processo cognitivo; e finalmente a mecânica ou elemento do jogo, que vai dizer como esse conteúdo será apresentado ou como se dará a interação com o mesmo.

Segundo Ferraz e Belhot [2010] a utilização do gerúndio do verbo, na taxonomia revisada, possibilita que seja esclarecido o “quê” e “como” avaliar, e se o objetivo foi ou não alcançado. Este verbo no gerúndio é o processo cognitivo, enquanto que o verbo no infinitivo, utilizado no início da frase é a categoria do qual o verbo no gerúndio faz parte Krathwohl [2002]. O “como” algo é aprendido no jogo sério pode ocorrer através do uso de uma mecânica desenvolvida para o jogo.

A dimensão dos processos cognitivos está dividida em 6 cartas, também baseadas nas definições apresentadas Krathwohl [2002] e Ferraz e Belhot [2010]. Cada carta representa um processo cognitivo, do menos complexo (carta “lembrar” o número “1”) ao mais complexo (carta “criar” com número “6”):

- **Lembrar:** Recuperar conhecimento relevante da memória de longo prazo - reconhecer ou recordar ideia ou conteúdo. Reconhecer requer distinguir e selecionar uma determinada informação. Recordar está relacionado à busca por uma informação relevante memorizada.
- **Entender:** Construir significado a partir de mensagens instrucionais, incluindo comunicação oral, escrita e gráfica. O aluno estabelece uma conexão entre o novo e o conhecimento previamente adquirido, além de conseguir reproduzir a informação com suas próprias palavras.
- **Aplicar:** Executar ou usar um procedimento em uma determinada situação. Pode também abordar a aplicação de um conhecimento numa situação nova.
- **Analisar:** Dividir a informação ou material em suas partes constituintes e detecção de como as partes se relacionam umas com as outras e com uma estrutura ou propósito geral.
- **Avaliar:** Relacionado a realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de eficiência e eficácia.
- **Criar:** significa juntar elementos para formar um todo novo e coerente; fazer um produto original; reorganizar elementos em um novo padrão ou estrutura; desenvolvimento de ideias novas e originais - ferraz e belholt; iowa state e Krathwohl

Cada uma das cartas possuem a definição do seu processo cognitivo, descrito acima, mais os verbos que serão utilizados.

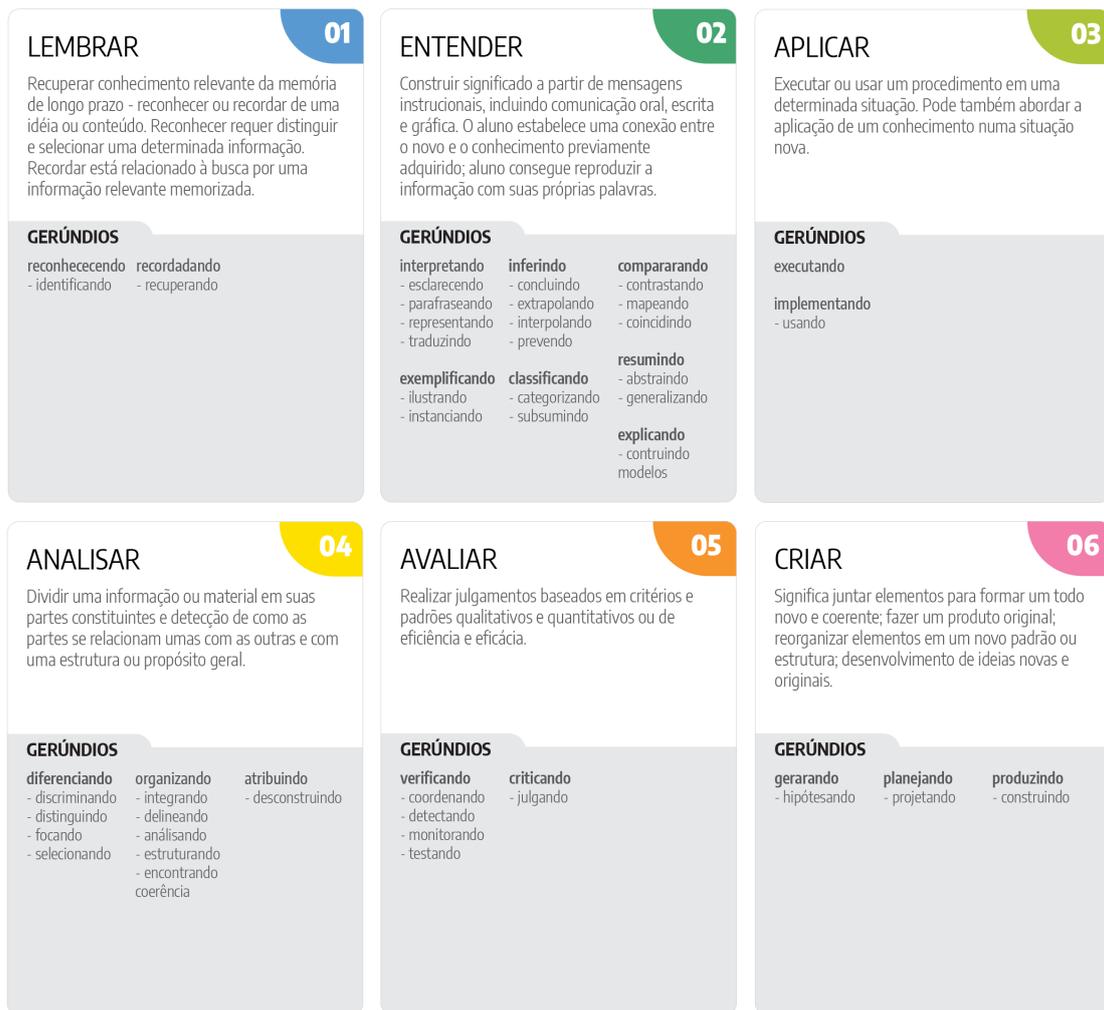


Figura 5.22. Cartas de Bloom da dimensão do processo cognitivo. Fonte: própria (2019)

No caso de uma avaliação do aprendizado obtido com o jogo, ela deve ser feita levando em consideração os objetivos educacionais e os níveis da Taxonomia Revisada de Bloom escolhidos para o jogo.

5.7.5. História

A história do jogo é a sequência de eventos que se desdobra no jogo Schell [2014]. Como foi explicado na seção sobre o modelo DPE de Winn [2009], existem dois tipos de história do jogo: a do designer e a do jogador. Neste ENDO-GDC o foco é a história projetada pelo designer, uma vez que a história do jogador só é passível de ser analisada após um *playtest*.

A história do designer é responsável por fornecer propósito, engajamento, definir cenários e transmitir o conteúdo e também é um meio para transmitir o conteúdo pedagógico.

A história é uma narrativa, ou uma proposta de narrativa que vai ser elaborada mais tarde. Para definir a história é necessário que existam personagens e um mundo. Para isso, a equipe deve procurar:

- Um momento histórico, real ou imaginário, no passado, presente ou futuro, ou um cenário, que define em que época e lugar a história acontece;
- Um, ou mais, personagens do jogador ou dos jogadores;
- Um, ou mais, personagens não jogadores (NPC, de *non-player character*, que são usados na narrativa para criar desafios ou fornecer informações, e
- um conflito principal, sobre o qual a narrativa vai ser construída.

Por exemplo, na família de jogos do Super Mário, o lugar é um mundo de fantasia, o personagem principal é o Mário, cuja missão, ou conflito, é salvar a princesa, capturada pelo Bauzer (ambos NPCs).

Jogos podem existir em algum lugar teórico em um amplo espectro entre o jogo totalmente abstrato, como *velha*, e um jogo totalmente linear, onde não há escolhas e decisões, como em um romance ou conto. Caberá ao grupo encontrar o ponto ideal do jogo, em função das necessidades do problema, dos jogadores e do conteúdo pedagógico.

A história não é obrigatória e pode ser substituída por apenas um tema, ou uma decoração das peças, ou mesmo um jogo totalmente abstrato, se o Conteúdo Pedagógico assim o permitir.

5.7.6. Feedback Educativo

Winn [2009] define *feedback* como conhecimento dos resultados que é crítico para dar suporte ao desempenho e a motivação do jogador. Para McGonigal e Rieche [2012] o *feedback* é importante, pois diz ao jogador o quão perto está de alcançar as metas do game. Dessa forma, ele é um elemento que está fortemente associado criação e manutenção da motivação do jogador, Winn [2009].

Em um jogo educativo o *feedback* vai auxiliar no processo de *debrieffing*, principalmente se houver uma ligação entre o conteúdo pedagógico e a mecânica de *feedback* do jogo. Assim é importante distinguir *feedbacks* relacionados a motivação e os relacionados

ao ensino.

O foco desta seção não é descrever as mecânicas de *feedback* – isso deve ser feito na seção de mecânicas do jogo, que será vista mais adiante. O objetivo aqui é chamar a atenção desse tipo de mecânica e tornar mais visíveis conceitos ou conhecimentos relacionados ao conteúdo pedagógico, que são importantes de serem representados ou estarem associados a uma mecânica de *feedback*. Podemos citar como exemplo o jogo do Sr. X: a cor de um mesmo personagem está relacionada ao sinal de um número e o desalinhamento de vagões está associados a ao desequilíbrio da equação de primeiro grau.

Então na seção de mecânicas do jogo dois tipos de mecânica de *feedback* devem aparecer: de motivação e os de educação. Este último deve estar relacionado a conceitos e conhecimentos listados aqui na seção de *Feedback Educativo*

Os *feedbacks* devem ser analisados durante o *playtest* para saber se estão auxiliando no processo de *debriefing* do jogador (revisão e análise dos eventos que aconteceram no jogo) e se este processo está auxiliando no aprendizado do conteúdo pedagógico descrito no ENDO-GDC

5.7.7. Objetivos do Jogo

O objetivo fornece um propósito ao jogador, ou seja, o que ele deve fazer para ganhar o jogo. Além disso, objetivos claros, específicos e difíceis levam ao aumento de desempenho do usuário [Locke e Latham, 1990].

De acordo com Garris, Ahlers e Driskell [2002] objetivos claros e específicos são cruciais no desencadeamento de maior atenção e motivação, pois permitem que o indivíduo perceba as discrepâncias entre o *feedback* que está recebendo e a meta/objetivo que deve alcançar. Dessa forma, se o *feedback* indica que o desempenho atual não atende aos objetivos estabelecidos, os usuários tentam reduzir essa discrepância. Quando há um alto comprometimento do jogador com os objetivos essa discrepância leva a um aumento no esforço e no desempenho [Kernan e Lord, 1990].

Na seção objetivos do jogo são descritos os objetivos que o jogador deve alcançar no jogo. Esses objetivos devem estar alinhados com a história do jogo e aos objetivos de aprendizado. Além disso, devem ser desenvolvidos *feedbacks* que informem ao jogador como está seu desempenho em relação a esses objetivos.

É comum que existam objetivos intermediários. Por exemplo, para ganhar em um jogo de quarteto, onde se colecionam cartas de 4 em 4, é preciso fazer uma coleção de cada vez. Montar um grupo completo é um objetivo intermediário. Ter o maior número de grupos completos é o objetivo final.

5.7.8. Modelo MDA

No centro do ENDO-GDC ficam, um bloco onde o conceito do jogo propriamente dito vai ser definido, de acordo com o modelo MDA. Todo o resto do ENDO-GDC serve de limitador e habilitador desse modelo.

Essa área pode ser usada para colocar os *Post-its* em separado, ou, com o uso adicional de tachas e fios, para construir diagramas semelhantes ao da Figura 5.8. De

qualquer forma, é importante que haja uma ligação entre as dinâmicas e a estética do jogo, e entre as dinâmicas e as mecânicas.

Para obter uma ontologia de estéticas, dinâmicas e mecânicas, o leitor pode se referir a J. S. Kritz [2020], J. Kritz, E. Mangeli e Xexeo [2017] ou Engelstein e Shalev [2020], ou ainda ao *site* Board Game Geek⁷.

5.7.8.1. Mecânicas

As mecânicas são cruciais para o *gameplay*, por isso são o núcleo, a essência de um jogo, e o que os distingue de outras mídias de entretenimento, como livros e filmes [Schell, 2014]. Elas estabelecem os procedimentos e as regras do jogo que definem o comportamento do jogador face ao jogo, ou seja, informam como o jogador pode alcançar os objetivos do game e também definem como funciona o mundo do jogo.

Na seção mecânicas devem ser definidas as principais mecânicas do jogo, ou seja, as regras e procedimentos que regem o *gameplay* do jogador e o do mundo do *game*. É importante relacioná-las com os objetivos de aprendizado pensados para o jogo, de forma que possibilite o desenvolvimento de um jogo endógeno. Essa mistura e do conteúdo pedagógico com as mecânicas e outros elementos do jogo possibilita que o jogador aprenda através do *gameplay*.

No *playtest* é importante verificar se as mecânicas são utilizadas como esperado, se atingem o objetivo de aprendizado ao qual foram associadas, e se as mecânicas geram as dinâmicas e estéticas desejadas. Um item importante a verificar é a retórica procedural, que é a prática de usar processos de maneira persuasiva, para mudar a opinião de uma pessoa ou modificar suas ações, [Bogost, 2008].

Mecânicas podem ter significados quando ligadas a história. *Action Points*[Engelstein e Shalev, 2020], onde o jogador recebe um número de pontos de ação para gastar em seu turno, pode ser ligados, por exemplo, a horas gastas com a produção de um recurso no jogo. Isso pode ser usado em um jogo onde se ensina a revolução industrial, mostrando a mudança de vida do artesão, mas também em um jogo onde se exercita a adição e a subtração por meio de cartas de números ou de operações, que podem exigir diferentes quantidades de pontos de ação para construir uma equação.

Ao desenvolver as mecânicas, o designer deve se preocupar com a forma que elas serão usadas. A mecânica de correr pode ser usada tanto para a dinâmica de fugir quanto para a dinâmica de atacar. Contextualmente, isso pode levar a uma falha no processo de aprendizagem. A mecânica de voto, em um jogo que simula um parlamento, pode fazer com que jogadores votem em propostas anti-éticas apenas para evitar que um adversário ganhe.

Um exemplo real que pode ser citado é o da criação de um jogo pelos autores sobre corrupção, que obviamente foi feito para ensinar crianças a serem honestas. Após o primeiro *playtest*, porém, a avaliação da partida mostrou que era muito mais divertido ser corrupto, o que passaria uma mensagem totalmente errada. Aquela versão do jogo foi

⁷<https://boardgamegeek.com/>

então reprovada.

J. Kritz, E. Mangeli e Xexeo [2017] fornecem e detalham mecânicas para jogos de tabuleiro, algumas existentes em videogames, listadas a seguir:

1. algoritmos;
 - (a) ação;
 - (b) programação de ações;
 - (c) leilão;
 - (d) *chit pull* - sortear fase;
2. coleção;
 - (a) construção de conjunto;
 - (b) coleção de conjuntos;
3. especulação de commodities;
 - (a) posse de ações (*stocks*);
4. pegar em um conjunto, possivelmente de forma aleatória (*draft*)
 - (a) pegar ações de um conjunto de possibilidades;
 - (b) pegar cartas;
5. se expressar de alguma forma;
 - (a) atuar;
 - (b) interpretar;
 - (c) cantar;
 - (d) contar histórias;
6. jogos de azar;
 - (a) apostar;
 - (b) jogar carta;
7. votar
8. desenhar linhas;
 - (a) construir redes;
9. mover;
 - (a) movimento ponto a ponto
 - (b) movimento em por áreas;
 - i. movimento em grid;
10. construção de padrões;
11. pressionar a sua sorte;
12. sorteio aleatório;
 - (a) sortear item em conjunto, normalmente cartas;
 - (b) jogar dados;
13. escolher posição secretamente;
14. seleção de ação simultânea;
15. “tome isso”;
16. posicionamento de ladrilho (*tile*);
17. troca;
18. objetivos;
 - (a) controlar de área;
 - (b) cercar área;
 - (c) memória;
 - (d) reconhecer padrões;
 - (e) coleta e entrega;
 - (f) eliminação de jogadores;
 - (g) gerenciamento de recursos;
19. conjunto de regras;
20. balanceamento;
 - (a) assimetria;
 - i. poderes variáveis do jogador;
 - (b) pedra papel tesoura;
21. fase;
 - (a) impulso;
 - (b) controle de tempo;
 - (c) ordem de fase variável;
22. times;
 - (a) jogo cooperativo;
 - (b) parceria;
23. representação dos dados do jogo;
 - (a) componente;
 - (b) áreas;
 - (c) tabuleiro modular;
 - (d) papel e lápis;
 - (e) padrão;
 - (f) dado;
 - i. dado como marcador;
 - (g) *token*;
 - (h) ladrilho (*tile*);
24. recursos;
 - (a) carta de ação;
 - (b) ponto de ação, e
 - (c) pontos de vitória.

5.7.8.2. Dinâmicas

As dinâmicas descrevem o comportamento em tempo de execução da mecânica, ou seja, como elas serão usadas pelos jogadores para eles atingirem seus objetivos parciais e finais. Elas são a ponte entre a estética e a mecânica.

Voltamos a um exemplo. Em Super Mario Bros. o herói Mario pode pular em cima de um inimigo, como os *Goombas*, ou pular sobre eles, simplesmente evitando-os e seguindo em frente. A mecânica usada é a mesma, correr e pular, o uso, porém, é diferente. É possível, inclusive, que um jogador iniciante não perceba que é possível cair sobre o *Goomba*, destruindo-o, porque pular sobre os inimigos e seguir com o jogo é uma prática comum em outros jogos, como Pitfall!, onde é impossível destruir os predadores que atacam o herói. Pular em cima, ou destruir, e pular por cima, ou evitar, são dinâmicas diferentes para a mesma mecânica.

Da mesma maneira, as duas ações trazem ,ao jogador, experiências emocionais distintas. Ambas partes do instinto de sobrevivência[Dillon, 2010], gerado pelo medo inspirado pela figura do *Goomba*, mas acertar o *Goomba* provavelmente levará a uma pequena sensação de alegria ou excitação[Dillon, 2010]. Já pular sobre o *Goomba* provavelmente trará uma sensação de alívio, causada pela cessação do Medo.

Nesta seção devem ser descritas as dinâmicas do jogo. Uma referência para auxiliar no preenchimento da seção é a lista feita por Vahlo et al. [2017], que possui as 33 dinâmicas mais comuns de se encontrar nos jogos atuais. A dissertação de J. S. Kritz [2020] também traz dinâmicas para jogos de tabuleiro.

A seguir, listamos as 29 dinâmicas encontradas por J. S. Kritz [2020] para jogos de tabuleiro, sendo algumas aplicáveis a videogames.

1. Sacrifício: sacrificar uma peça ou posição para maiores ganhos;
2. Efeito indireto: executar algo ao querer uma consequência de seu eventos;
3. Obter informações: usar uma ação descobrir algo de interesse;
4. Opções de redução: usar uma ação para reduzir as opções de outros jogadores;
5. Extinção de recursos: reduzir a quantidade de um recurso limitado ou torná-lo inútil;
6. Dedução: usar informações abertas para descobrir informações escondidas;
7. Mudança de estado do jogo: usar uma ação para provocar propositalmente uma mudança em o jogo;
8. Combo: produzir efeitos automáticos em cadeia de o jogo;
9. Bloqueio: bloquear outro jogador por meio de uma ação, estratégia, ou progresso;
10. Um contra todos: quando um jogador ataca todos os outros simultaneamente;
11. Todos contra um: todos os jogadores do jogo, menos um, unem forças para derrotar o outro jogador;
12. Aliança: quando os jogadores unem forças para alcançar benefício mútuo;
13. Interpretação específica das regras: use um ponto de vista específico para criar melhores benefícios para si mesmo;
14. Objetivo próprio: perseguir um objetivo autoproclamado diferente do objetivo do jogo;
15. Jogue com segurança: não correr riscos, jogar apenas na certeza;

16. Jogo de risco: aceitar riscos maiores buscando maiores recompensas;
17. Sobrevivência: jogar apenas para evitar a eliminação;
18. Acampar: ficar em uma posição ou ação por muito tempo;
19. Protecionismo: proteger uma posição ou peças;
20. Planejamento de ação: jogar de acordo para suas próximas ações, planeje uma série de ações;
21. Apresse o jogo: acelerar o final do jogo;
22. Estratégia flexível: mudar a estratégia por causa do estado do jogo;
23. Rejeitar objetivos: intencionalmente não alcançar o objetivo de um jogo para obter alguma vantagem;
24. Intimidar: usar de uma posição mais forte para forçar outro jogador a jogar como você quiser;
25. Distração: usar uma ação para mude a atenção dos outros jogadores de sua real intenção ou objetivo;
26. Conversa fiada: falar o tempo todo para distrair outros jogadores;
27. Contar recursos: usar o conhecimento prévio dos recursos disponíveis para contá-los e obter vantagem;
28. Blefe: transmitir informações falsas para manipular as ações de outros jogadores, e;
29. Convencer: convencer outros jogadores.

5.7.8.3. Estética

As estéticas descrevem as respostas emocionais que deseja-se evocar no jogador [Hunicke, Leblanc e Zubek, 2004], quando ele interage com o sistema do jogo e os elementos que dão aparência sensorial ao game [Schell, 2014]. O sentido múltiplo do nome da seção é importante, pois as emoções podem ser parte central de um jogo educativos (principalmente se um dos resultados de aprendizado for o afetivo), assim como o visual pode passar informação ao jogador, sobre como algo é representado em termos de forma e cores, que podem também podem ajudar a intensificar uma emoção do jogador.

Na seção Estética devem ser definidas as emoções que se deseja evocar no jogador, assim como deve-se caracterizar o design geral do jogo. Algumas sugestões de estética são feitas no artigo de Hunicke, Leblanc e Zubek [2004], mas Dillon [2010] propõe o modelo 6-11 para tratar exatamente desse ponto.

Usar o modelo 6-11 permite à equipe criando o jogo desenhar um modelo como o da figura 5.8, que traz *insights* bastante interessantes ao trabalho.

Uma opção adicional é substituir alguns ou todas as emoções e os instintos propostas por Dillon [2010] pelas emoções, experiências e respostas propostas por Ekman [2021]⁸.

Optando pelo “*Atlas of Emotions*”, pode ser usado um modelo mais sofisticado de compreensão das emoções, que inclui pré-condições, eventos que geram o estado emocional a partir de uma percepção pessoal, mudanças físicas e psicológicas causadas pelo

⁸<http://atlasofemotions.org/>

estado emocional, ações, construtivas e destrutivas, tudo isso levando a uma pós condição Ekman [2021].

No caso de jogos educacionais, alguns cuidados adicionais, porém, devem ser tomados em relação as estéticas propostas, já que certas emoções podem não ser adequadas. Por exemplo, na maioria dos casos, não é desejado estimular a agressividade em sala de aula, no ensino formal, porém no caso de instruções militares, isso pode ser necessário, e talvez a tristeza tenha que ser evitada. Um caso interessante foi proposto a um dos autores, relacionado ao ensino militar ligado a tomada de decisões sob pressão: o aluno, nesse caso, não pode sucumbir as emoções, e, dependendo do tipo de objetivo educacional, isso significa que o jogo deve ou não provocar emoções fortes, já que em alguns momentos se deseja, e em outros não, levar o aluno a buscar a decisão racional mesmo em frente a fortes emoções.

Outro caso interessante que os autores tiveram a oportunidade de trabalhar estava ligado a emoção medo. Em um jogo de educação sexual, deveria ser gerada uma consciência adequada dos riscos da falta de prevenção, mas não ao ponto de dar aos alunos uma sensação de medo em relação à prática do sexo seguro.

5.7.9. Inspirações

O preenchimento da seção inspirações deve vir em qualquer momento a partir da definição das seções do bloco de público-alvo e aprendizado. O objetivo dessa seção é reunir referência de jogos para o *game* que será desenvolvido. Da mesma forma como ocorre com outros produtos, pesquisar sobre jogos similares pode mostrar o que já deu certo ou errado, para o jogo do mesmo tipo ou com a mesma temática e também pode ajudar a construir diferenças para o projeto.

Ao utilizar as lentes do design holográfico [Schell, 2014], as seguintes perguntas devem se para os jogos referenciados:

- Quais elementos do jogo tornam a sua experiência agradável?
- Que elementos do jogo tornam sua experiência desagradável?
- Como posso mudar a os elementos do jogo para melhorar sua experiência?

5.7.10. Restrições

Tendo em vista que visamos um projeto conceitual, queremos deixar em aberto ainda as opções de tecnologia ou plataforma, a serem discutidas a seguir.

Porém, mesmo durante o projeto conceitual, podem surgir demandas que restringem as opções. Por isso, após uma evolução do ENDO-GDC original [Taucei, 2019] onde apareciam tecnologia e plataforma, decidimos unificá-las no quadro restrições.

5.7.10.1. Tecnologias

Um dos últimos elementos a serem definidos, tanto no modelo DPE de Winn [2009] e no modelo da tétrede elementar de Schell [2014] é a tecnologia. Segundo o autor deixar a escolha da tecnologia como um dos últimos elementos é importante, pois "...é fácil se deixar levar pelas possibilidades do que a tecnologia pode fazer e esquecer que o propósito

é criar um ótimo jogo”[Schell, 2014].

A tecnologia é o meio que será utilizado para desenvolver o jogo e que vai ajudar a entregar a experiência que deseja-se que o jogador tenha com o *game*. A sua escolha vai definir as restrições e os recursos que serão necessários para projeto. Os recursos são os instrumentos que serão utilizados para desenvolver o jogos as restrições são as limitações impostas pelos instrumentos e pelas outras camadas ao jogo.

5.7.10.2. Plataforma

Meio onde o jogo irá rodar. Pode ser:

- **Analógico:** como jogos de cartas ou jogos de tabuleiro
- **Digital:** celulares (apple ou android), PC (steam, origin,); consoles (Xbox, Playstation, Nintendo)

Cada plataforma selecionada para o jogo rodar deve ser escrita em post-it diferente. Em muitos casos a definição da plataforma também acaba definindo onde o jogo será comercializado.

5.8. Exemplo de uso do ENDO-GDC

O Mapa do Tesouro é um jogo que tem o propósito de ensinar pensamento computacional concorrente a crianças de 4 a 9 anos. O desenvolvimento do jogo se realizou por um método iterativo e incremental, e o ENDO-GDC não foi usado desde o início, então alguns comentários mostram *insights* que o primeiro uso do ENDO-GDC gerou, a partir da observação do jogo e de entrevistas com o criador principal.

O jogo procura resolver os seguintes problemas:

- o pensamento computacional concorrente é uma necessidade do futuro, e
- o aprendizado por meio de linguagens de programação é muito complexo para a idade alvo.

Na seção Jogadores/Alunos do jogo foi listado o seguinte perfil:

- crianças de 4 à 9 anos.

O conteúdo pedagógico inclui:

1. Pensamento Computacional
2. Pensamento Computacional Concorrente
3. Sistemas simbólicos e Representações
4. Eficiência e Desempenho
5. Decomposição de Problema
6. Resolução de Problemas
7. Desenvolvimento de Algoritmos
8. Exclusão Mútua

Após listar o conteúdo pedagógico abordado, foi possível perceber melhorias poderiam ser feitas no *game*: o número de blocos de comandos utilizados na fase poderia estar sempre visível, mostrando o número máximo de blocos permitidos; as fases mais

complexas poderiam explicitar ou sugerir formas de decompor o problema, para ensinar o jogador a dividi-lo em problemas menores e mais fáceis de resolver. Isso reforçaria a presença dos assuntos, que já estavam presentes, mas não eram devidamente tratados.

Os objetivos de aprendizado foram:

- APLICAR Sistemas simbólicos e Representações, USANDO as mecânicas de andar, cavar, levantar bandeira, esperar e atirar no jogo
- APLICAR o desenvolvimento de algoritmos EXECUTANDO as mecânicas de andar, cavar, levantar bandeira, esperar e atirar para levar o pirata até o tesouro

A história do jogo era do tipo tema e não tinha continua elementos do conteúdo pedagógico sendo referenciados:

- dois piratas vão parar em uma ilha em busca de tesouro.

O Jogo possui os seguintes objetivos:

1. achar o tesouro,
2. achar o diamante (opcional),
3. usar poucos blocos, e
4. completar todas as fases.

Muitas das mecânicas presentes no jogo estavam na forma de peças de comando. Através dessas peças, o jogador programava os passos executados pelo pirata, como pode ser visto na figura 5.4, à esquerda. As seguintes mecânicas foram encontradas:

- ações
 - olhar para uma direção;
 - andar;
 - cavar;
 - levantar bandeira;
 - esperar;
 - atirar;
 - empurrar barril, e
- programação de ações.

Esse jogo, por sua construção, tem mecânicas muito próximo da dinâmicas desejadas, porém ainda forma identificadas as seguintes mudanças:

- efeito indireto, quando se atira em ou empurra algo;
- mudança de estado;
- planejamento de ação;

Ao criar o jogo foi usado o modelo 6-11. As estéticas que foram buscadas foram:

- apreciação da cor;
- ganância e coleta, para obter os tesouros e diamantes;
- alegria, por completar uma fase, e
- orgulho por completar o jogo

Os *feedbacks* educacionais, principalmente aqueles relacionados ao desempenho do jogador, eram apresentados somente no final da fase (Figura 5.23):



Figura 5.23. Imagem do jogo Mapa do Tesouro, com o resultado da fase. Fonte: imagem retirada de uma sessão do jogo Mapa do Tesouro Nascimento, G. B. Xexéo e Brito [2019]

- total de passos realizados pelo pirata, que indica a qualidade de uma resposta certa;
- tesouros e diamantes encontrados, que indica se o jogador fez o problema mais completo;
- erro sinalizado com um "X" na cabeça do pirata, que indica que o jogador não completou a fase, logo não usou os conceitos de forma correta, e
- indicador de fase, que indica que o jogador está aprendendo.

As restrições foram as plataformas alvo, Android e PC; e a tecnologia, Unity.

O resultado aparece na Figura 5.24.

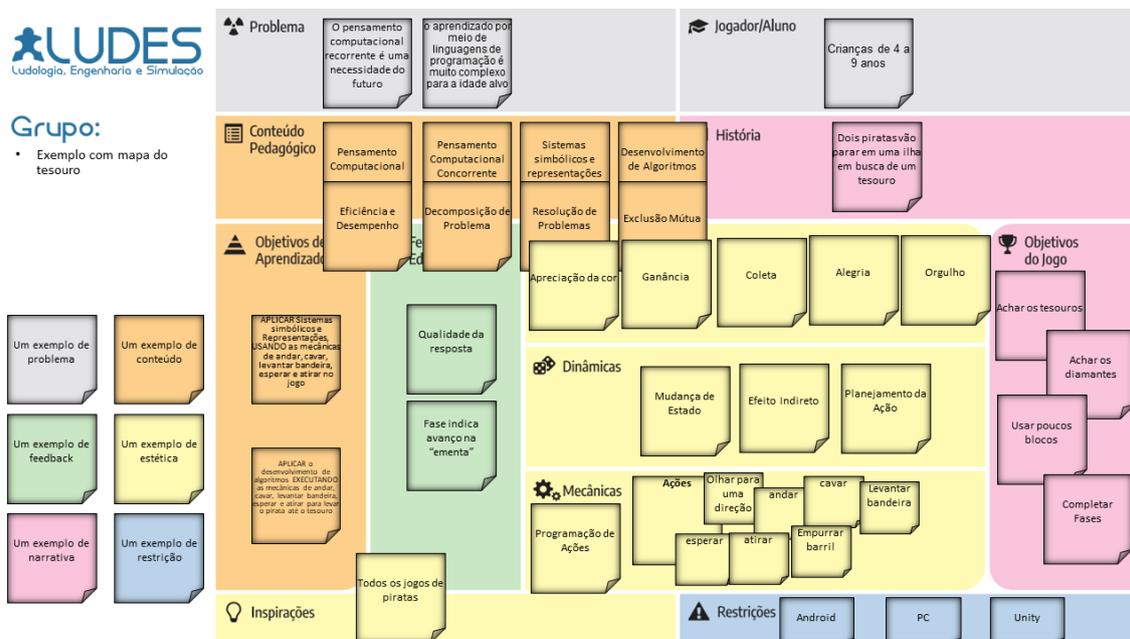


Figura 5.24. Exemplo do ENDO-GDC preenchido para o jogo Mapa do Tesouro. Fonte: própria

5.9. Conclusões

Taucei, L. Costa e G. Xexéo [2018] avaliaram vários canvas para jogos, e mais tarde, Taucei [2019] construiu o ENDO-GDC, a partir da análise das propriedades importantes de um canvas e das necessidades do desenvolvimento de jogos educacionais endógenos. Nesse mesmo trabalho, mostrou que o ENDO-GDC também pode ser utilizado para fazer uma análise de um jogo já existente.

Neste tutorial, apresentamos o ENDO-GDC, e estendemos ainda mais seu uso, indicando como algumas técnicas típicas de criatividade, como *user personas*[Tracy, 2015], podem ser usadas para auxiliar o desenvolvimento do canvas.

Também procuramos detalhar mais algumas fases, inclusive mostrando exemplos de *post-its* que podem aparecer nos quadros internos do ENDO-GDC.

Buscamos também apresentar mais justificativas para seu uso.

Este capítulo, em sua totalidade, busca não só levar o leitor a entender como o ENDO-GDC funciona, mas também estar apto a utilizá-lo em sua prática do desenvolvimento de jogos educacionais.

Com o resultado, o processo de criação do jogo pode seguir vários caminhos, mas espera-se que sejam sempre ligados a um processo iterativo e incremental. Muito do resultado de uma iteração completa do jogo pode ser levado para um documento comumente citado na literatura, mas fracamente definido e sempre em evolução, o *Game Design Document* ou o *Game Bible*.

Agradecimentos

Nossos agradecimentos especiais a Marcus Parreira e Alexandre Uchoa, que revisaram este texto com bastante cuidado, e nossas desculpas por não conseguir atender a todos os seus pedidos.

Links importantes

Todo material disponível sobre o ENDO-GDC pode ser encontrado em <https://github.com/LUDES-PESC/ENDO-GDC> sob a Licença MIT.

A página a seguir contém uma versão em A4 do ENDO-GDC. No repositório pode ser encontrada uma versão em 40.0 por 27.8 cm.

 Problema

 Jogador/Aluno

 Conteúdo Pedagógico

 História

 Objetivos de Aprendizado

 Feedbacks Educativos

 Estética

 Objetivos do Jogo

 Dinâmicas

 Mecânicas

 Inspirações

 Restrições

Referências

- Abdul Jabbar, A. I. e P. Felicia (2015). “Gameplay Engagement and Learning in Game-Based Learning”. Em: *Review of Educational Research* 85.4, pp. 740–779.
- Abt, Clark C. (mar. de 2002). *Serious Games*. English. Lanham, MD: UPA. ISBN: 978-0-8191-6148-2.
- Barbosa, Eduardo Fernandes (s.d.). *Gerência de Qualidade Total na Educação*. Belo Horizonte.
- Bates, Bob (2004). *Game Design*. 2ª ed. Boston, MA: Thomson.
- Bishop, Jacob Lowell e Matthew A. Verleger (2013). “The flipped classroom: A survey of the research”. Em: *ASEE national conference proceedings, Atlanta, GA*. Vol. 30. Issue: 9, pp. 1–18.
- Bogost, Ian (2008). “The rhetoric of video games”. Em: *The ecology of games: Connecting youth, games, and learning*, pp. 117–140.
- BRASIL (2018). *Base Nacional Comum Curricular*. Brasília: Ministério da Educação.
- Brown, Dan M. (2010). *Communicating Design: Developing Web Site Documentation for Design and Planning (2nd Edition) (Voices That Matter)*. 2ª ed. Berkeley, CA: New Riders. ISBN: 0321712463.
- Callois, Roger (2001). *Man, Play and Games*. Illinois: University of Illinois Press.
- Carey, Richard (abr. de 2015). *Game Design Canvas – Richard Carey Digital Media*. URL: <http://richardcarey.net/game-design-canvas/> (acesso em 25/04/2018).
- CEDEFOP (2014). *Terminology of European education and training policy: A selection of 130 key terms*. 2ª ed. Luxembourg: European Centre for the Development of Vocational Training, Publications Office of the European Union.
- Costa, Leandro Demenciano (2008). “O que os Jogos de Entretenimento Têm que os Jogos com Fins Pedagógicos Não Têm: Princípios para Projetos de Jogos com Fins Pedagógicos”. Tese de dout. Rio de Janeiro: PUC/RJ.
- Crawford, Chris (28 de jun. de 2003). *Chris Crawford on Game Design*. New Riders Games. ISBN: 0-13-146099-4.
- Dillon, R. (2010). *On the Way to Fun: An Emotion-Based Approach to Successful Game Design*. Natick, Massachusetts: A K Peters Ltd.
- Dzgoeva, Svetlana (30 de mar. de 2016). “Development of a Gamification Design Method for a Business Process Modeling Tool”. Diss. de mest. Münster: Department of Information Systems, Westfälische Wilhelms-Universität Münster.
- Ekman, Paul (2021). *Atlas of Emotions*. URL: <http://atlasofemotions.org/> (acesso em 29/08/2021).
- Engelstein, Geoffrey e Isaac Shalev (2020). *Building Blocks of Tabletop Game Design: An Encyclopedia of Mechanisms*. Boca Raton, FL: CRC PRESS.
- Ferraz, Ana Paula do Carmo Marcheti e Renato Vairo Belhot (2010). “Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais”. Em: *Gestão & Produção* 17.2, pp. 421–431.
- Gagne, Robert M. et al. (jun. de 2004). *Principles of Instructional Design*. English. 5 edition. Belmont, CA: Cengage Learning. ISBN: 978-0-534-58284-5.
- Garris, Rosemary, Robert Ahlers e James E. Driskell (dez. de 2002). “Games, Motivation, and Learning: A Research and Practice Model”. en. Em: *Simulation & Gaming*

- 33.4, pp. 441–467. ISSN: 1046-8781. DOI: [10.1177/1046878102238607](https://doi.org/10.1177/1046878102238607). URL: <https://doi.org/10.1177/1046878102238607> (acesso em 07/12/2017).
- Gray, Dave, Sunni Brown e James Macanufo (2010). *Gamestorming*. Sebastopol, CA: O'Reilly.
- Halverson, Richard (2005). “What Can K-12 School Leaders Learn from Video Games and Gaming?” en. Em: *Innovate: Journal of Online Education* 1.6. ISSN: 1552-3233. (Acesso em 30/04/2019).
- Huizinga, Johan (1938–2001). *Homo Ludens: O Jogo Como Elemento da Cultura*. Trad. por João Paulo Monteiro. 5ª ed. São Paulo (SP): Perspectiva. ISBN: 978-85-273-0075-9.
- Hunicke, Robin, Marc Leblanc e Robert Zubek (2004). “MDA: A Formal Approach to Game Design and Game Research”. Em: *Nineteenth National Conference of Artificial Intelligence*. San Jose, California, USA.
- Järvinen, Aki (2008). “Games without frontiers: Theories and methods for game studies and design”. Tese de dout. Finland: University of Tampere.
- Jiménez, Sergio (2013). *Gamification Model Canvas - Gamasutra*. en. URL: https://www.gamasutra.com/blogs/SergioJimenez/20131106/204134/Gamification_Model_Canvas.php (acesso em 08/06/2018).
- Juul, Jesper (mar. de 2012). *A Casual Revolution – Reinventing Video Games and Their Players*. Inglês. Cambridge, Mass.; London: MIT Press. ISBN: 978-0-262-51739-3.
- Kalmpourtzis, G. (2018). *Educational Game Design Fundamentals: A Journey to Creating Intrinsically Motivating Learning Experiences*. 1st. A K Peters/CRC Press.
- Kernan, Mary C e Robert G Lord (1990). “Effects of valence, expectancies, and goal-performance discrepancies in single and multiple goal environments.” Em: *Journal of applied psychology* 75.2, p. 194.
- Koster, Raph (nov. de 2004). *Theory of Fun for Game Design*. Paraglyph Press. ISBN: 1-932111-97-2.
- Krathwohl, David R. (nov. de 2002). “A Revision of Bloom’s Taxonomy: An Overview”. Em: *Theory Into Practice* 41.4, pp. 212–218. ISSN: 0040-5841. DOI: [10.1207/s15430421tip4104_2](https://doi.org/10.1207/s15430421tip4104_2). URL: https://doi.org/10.1207/s15430421tip4104_2 (acesso em 14/06/2019).
- Kritz, J., E. Mangeli e G. Xexeo (2017). “Building an ontology of boardgame mechanics based on the boardgamegeek database and the mda framework”. Em: *Proceedings of SBGames 2017*. Curitiba, PR.
- Kritz, Joshua Silveira (2020). “An Ontology of Board Games Based On The MDA Framework / Uma Ontologia de Jogos de Tabuleiro Fundamentada no Framework MDA”. Diss. de mestr. Rio de Janeiro: Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ.
- Lam, Budd Royce (2015). *Game Design Canvas – Budd Royce Lam*. URL: <http://www.buddroyce.com/index.php/tools/game-design-canvas/> (acesso em 10/11/2017).
- Larmer, John, John Mergendoller e Suzie Boss (2015). *Setting the Standard for Project Based Learning: A Proven Approach to Rigorous Classroom Instruction*. Alexandria, VA: ASCD.

- Lieberman, Debra A. (2006). “What can we learn from playing interactive games”. Em: *Playing video games: Motives, responses, and consequences*, pp. 379–397.
- Locke, Edwin A. e Gary P. Latham (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.
- Marklund, Björn Berg (2015). “Unpacking Digital Game-Based Learning: The complexities of developing and using educational games”. Tese de dout. Skövde: University of Skövde.
- Marklund, Björn Berg et al. (14 de set. de 2019). “What Empirically Based Research Tells Us About Game Development”. Em: *The Computer Games Journal*.
- McGonigal, Jane (2011). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin Group, The. ISBN: 1594202850.
- McGonigal, Jane e Eduardo Rieche (ago. de 2012). *A realidade em jogo: Por que os games nos tornam melhor e como eles podem mudar o mundo: Por que os games nos tornam melhor e como eles podem mudar o mundo*. Português. Best Seller. ISBN: 978-85-7684-522-5.
- Michael, David R. e Sande Chen (2006). *Serious Games: Games that Educate, Train and Inform*. en. Thomson Course Technology. ISBN: 978-1-59200-622-9.
- Moore, Michael E. (2011). *Basics of Game Design*. Boca Raton, FL: CRC Press.
- Nascimento, Pedro, Geraldo Bonorino Xexéo e Eduardo Freitas Mangeli de Brito (2019). “Metodologias e Técnicas para Criação e Avaliação de Jogos Educacionais”. Em: 41ª Jornada Giulio Massarini de Iniciação Científica, 10ª SIAC - Semana de Integração Acadêmica da UFRJ. Rio de Janeiro. URL: <https://sistemasiac.ufrj.br/cadernoController/gerarCadernoResumo/36000000> (acesso em 29/08/2021).
- Newzoo (jun. de 2021). *2021 Global Games Market report*. URL: <https://newzoo.com/key-numbers/> (acesso em 21/08/2021).
- O’Hagan, Ann Osborne, Gerry Coleman e Rory V. O’Connor (jun. de 2014). “Software Development Processes for Games: A Systematic Literature Review”. Em: *21th European Conference on Systems, Software and Services Process Improvement (EuroSPI 2014)*. Vol. 425. CCIS. Springer-Verlag, pp. 182–193.
- Oliveira, Maria Victória (9 de nov. de 2015). *Jogo introduz aprendizado da álgebra de forma divertida*. Porvir - Inovações em Educação. URL: <https://porvir.org/jogo-introduz-aprendizado-da-algebra-de-forma-divertida/> (acesso em 29/08/2021).
- OMG (jan. de 2013). *Business Process Model and Notation (BPMN) Version 2.0.2*. Standard. Object Management Group. URL: <http://www.omg.org/spec/BPMN/2.0>.
- Osterwalder, Alexander (2004). “The business model ontology a proposition in a design science approach”. Tese de dout. Université de Lausanne, Faculté des hautes études commerciales.
- Osterwalder, Alexander e Yves Pigneur (jul. de 2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, NJ: John Wiley e Sons.
- Parry, Scott. B. (jul. de 1996). “The quest for competencies”. Em: *Training* 33, pp. 48–54.
- Plass, Jan L., Richard E. Meyer e Bruce D. Homer, ed. (2019). *Handbook of Game-Based Learning*. Cambridge, MA: The MIT Press.

- Prado, L. L. (jul. de 2018). “Jogos de Tabuleiro Modernos como Ferramenta Pedagógica: Pandemic e o Ensino de Ciências”. Em: *Revista Eletrônica Ludus Scientiae (ReLuS)* 2.2, pp. 26–38.
- Prensky, Marc (2001). “Fun, Play and Games: What Makes Games Engaging”. Em: *Digital Game-Based Learning*. McGraw-Hill. Cap. 5.
- (out. de 2003). “Digital Game-Based Learning”. Em: *Comput. Entertain.* 1.1, p. 21. DOI: [10.1145/950566.950596](https://doi.org/10.1145/950566.950596). URL: <https://doi.org/10.1145/950566.950596>.
- Pressman, Roger S. e Bruce Maxim (2019). *Software Engineering: A Practitioner’s Approach*. 9ª ed. New York, NY: McGraw-Hill. 976 pp. ISBN: 978-1-260-54800-6.
- Rouse, Richard (mar. de 2010). *Game Design: Theory and Practice, Second Edition*. en. Jones & Bartlett Learning. ISBN: 978-1-4496-3345-5.
- S. B. Júnior, José (2020). “Jogos Digitais Educacionais: Uma Revisão Sistemática da Literatura.” Tese de dout. Catalão – GO: Unidade Acadêmica Especial de Biotecnologia, Universidade Federal de Goiás – Regional Catalão.
- Salen, Katie e Eric Zimmerman (set. de 2003). *Rules of Play: Game Design Fundamentals*. Inglês. Cambridge, Mass: Mit Press. ISBN: 978-0-262-24045-1.
- Schell, Jesse (nov. de 2014). *The Art of Game Design: A book of Lenses*. English. 2ª ed. Boca Raton: A K Peters/CRC Press.
- Schonhaler, Frank et al. (2012). *Business Processes for Business Communities: Modeling Languages, Methods, Tools*. New York: Springer.
- Skinner, Ellen e Michael J. Belmont (dez. de 1993). “Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement Across the School Year”. Em: *Journal of Educational Psychology* 85, pp. 571–581. DOI: [10.1037/0022-0663.85.4.571](https://doi.org/10.1037/0022-0663.85.4.571).
- Squire, Kurt e Henry Jenkins (2003). “Harnessing the power of games in education”. Em: *Insight* 3.1, pp. 5–33.
- Suits, B. e T. Hurka (2014). *The Grasshopper - Third Edition: Games, Life and Utopia*. 3ª ed. Peterborough, Ontario: Broadview Press.
- Susi, Tarja, Mikael Johannesson e Per Backlund (nov. de 2007). *Serious games: An overview*.
- Sutton-Smith, Brian (1997). *The Ambiguity of Play*. Cambridge, Massachusetts: Harvard University Press. ISBN: 9780674017337.
- Taucei, Bernardo (2019). “ENDO-GDC: Desenvolvimento de um Game Design Canvas para Concepção de Jogos Educativos Endógenos”. Diss. de mestr. Rio de Janeiro, Brasil: Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ.
- Taucei, Bernardo, Luis Costa e Geraldo Xexéo (2018). “Uma análise crítica sobre canvas para jogos, baseado nas qualidades do Business Model Canvas e Design Thinking Canvas”. Em: *SBC – Proceedings of SBGames 2018 — Art & Design Track – Short Papers*.
- Tracy, Brian (2015). *Creativity & Problem Solving*. New York: AMACOM.
- Trevisan, André Luis e Roseli Gall do Amaral (jun. de 2016). “A Taxionomia revisada de Bloom aplicada à avaliação: um estudo de provas escritas de Matemática”. pt. Em: *Ciência & Educação (Bauru)* 22.2, pp. 451–464. (Acesso em 08/09/2019).

- Tsutsumi, Myenne Mieko Ayres et al. (jun. de 2020). “Avaliação de jogos educativos no ensino de conteúdos acadêmicos: Uma revisão sistemática da literatura”. Em: *Revista Portuguesa de Educação* 33.1, pp. 38–55.
- Vahlo, Jukka et al. (mar. de 2017). “Digital Game Dynamics Preferences and Player Types”. Em: *Journal of Computer-Mediated Communication* 22.2, pp. 88–103. ISSN: 1083-6101. DOI: [10.1111/jcc4.12181](https://doi.org/10.1111/jcc4.12181), eprint: <http://oup.prod.sis.lan/jcmc/article-pdf/22/2/88/19946828/jjcmcom0088.pdf>. URL: <https://doi.org/10.1111/jcc4.12181>.
- Van Eck, Richard (jan. de 2006). “Digital Game Based LEARNING It’s Not Just the Digital Natives Who Are Restless”. Em: *EDUCAUSE* 41.
- Verified Market Research (2020). *Educational Games Market Size And Forecast*. URL: <https://www.verifiedmarketresearch.com/product/educational-games-market/> (acesso em 31/08/2021).
- von Scheel, Henrik et al. (2015). “Phase 1: Process Concept Evolution”. Em: *The Complete Business Process Handbook: Body of Knowledge from Process Modelling to BPM*. Ed. por Mark von Rosing, August-Wilhelm Scheer e Henrik von Scheel. Vol. 1. Boston: Morgan Kaufmann, pp. 1–9.
- Winn, Brian M. (2009). “The design, play, and experience framework”. Em: *Handbook of research on effective electronic gaming in education*. IGI Global, pp. 1010–1024.
- Xexéo, Geraldo, A Carmo et al. (2017). *O Que São Jogos?* Technical report ES-752/17. Rio de Janeiro: LUDES - Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ. URL: <https://www.cos.ufrj.br/index.php/pt-BR/publicacoes-pesquisa/details/15/2766> (acesso em 07/06/2021).
- Xexéo, Geraldo, Eduardo Mangeli et al. (2021). “Games as Information Systems”. Em: *XVII Brazilian Symposium on Information Systems*. SBSI 2021. Uberlândia, Brazil: Association for Computing Machinery. ISBN: 9781450384919. URL: <https://doi.org/10.1145/3466933.3466961>.
- Zabala, Antoni e Laia Arnau (2014). *Como aprender e ensinar competências*. Porto Alegre: Artmed.

Biografias

Geraldo Xexéo, D.Sc. em Engenharia de Sistemas e Computação (COPPE-UFRJ/1995) e engenheiro eletrônico (IME/1988), é professor da Universidade Federal do Rio de Janeiro, onde coordena o LUDES - Laboratório de Ludologia, Engenharia e Simulação. Orientou a criação do ENDO-GDC. Lattes <http://lattes.cnpq.br/4783565791787812>

Bernardo Taucei, mestre em Engenharia de Sistemas e Computação (COPPE-UFRJ/2019) e Bacharel em Ciência da Computação (DCC-IM-UFRJ/2016), é designer. Criou o ENDO-GDC. Lattes <http://lattes.cnpq.br/7811304158988102>

Chapter

6

Pensando-fazendo Sistemas de Informação com Ética. Da pesquisa à engenharia, e vice-versa.

Luiz Paulo Carvalho, Flávia Maria Santoro, Rosa Maria M. Costa, Jonice Oliveira

Abstract

Our goal is to address concepts and definitions of Ethics and how they apply to Information Systems, both from a practical and research point of view. For this, the topics of ethics applied to research in Information Systems and ethics applied to Information Systems engineering will be conducted. Real cases and current and future perspectives and interaction are presented and discussed. We cover the relationship between Ethics and Information Systems, including recurring and essential topics in both thinking and making information systems, data protection and privacy, cyber crimes, intellectual property, professional ethics, plagiarism, consent, ethics committee, among others.

Resumo

Nosso objetivo consiste em abordar conceitos e definições de Ética e como se aplicam à Sistemas de Informação, tanto do ponto de vista prático como da pesquisa. Para isso, serão conduzidos os temas de ética aplicada à pesquisa em Sistemas de Informação e ética aplicada à engenharia de Sistemas de Informação. São apresentados e discutidos casos reais e perspectivas atuais e futuras e interação. Cobrimos a relação entre a Ética e Sistemas de Informação, incluindo tópicos recorrentes e essenciais tanto ao pensar quanto ao fazer sistemas de informação, proteção de dados e privacidade, crimes cibernéticos, propriedade intelectual, ética profissional, plágio, consentimento, comitê de ética, entre outros.

6.1. Introdução

Ética é um tema oriundo da Filosofia pouco abordado nos melhores cursos de Computação brasileiros, como oficialmente categorizados e classificados no Brasil pelo Ministério da

Educação e Cultura (MEC) [Carvalho et al. 2021a]. Como a área que analisa as práticas humanas, assim como suas consequências, podemos fazer diversos atravessamentos associativos à prática da engenharia de soluções computacionais e à prática da pesquisa em Computação. A ausência de consciência ou base Ética na formação da proficiência em Computação, seja formal [Carvalho et al. 2021a] ou informal [Carvalho et al. 2021b], aliena os praticantes da Computação sobre o pensa-fazer associado aos valores, diretos ou indiretos, de suas ações.

Neste trabalho buscamos fugir de uma limitação estruturalista e seguimos duas linhas, o pensar em Computação, tradicionalmente nomeada “estudo” e o fazer Computação, tradicionalmente nomeada “design” ou “construção”. Neste sentido, consideramos a opção pensar-fazer (ou fazer-pensar), combinando as percepções de ambas as práticas. Cabe uma ressalva, que também será reforçada nas Seções 6.3 e 6.4, ao tratarmos de pesquisar e engenheirar¹ há tanto formalização quanto adequação às instituições sociais associadas a estas práticas. Para que um estudo seja considerado uma pesquisa, e um design ou construção seja considerado uma engenharia há necessidade de metodologias e métodos bem-estruturados e estabelecidos às suas epistemologias, conceituais ou práticas. Estão alicerçadas sobre certo rigor paradigmático, metodológico e metódico. Isto não desqualifica ou diminui a qualidade de estudos, designs ou construções informais, cada um tem a sua pertinência e importância.

Há uma diferença entre Sistemas de Informação (SI) enquanto área ou disciplina de conhecimento, e sistema de informação enquanto um tipo de sistema envolvendo dados e informações. O primeiro tem propriedade de elemento único, como ramo científico; o segundo é uma instância de categoria estruturada e delimitada de sistema, i.e., há diversos sistemas de informação enquanto soluções computacionais enquanto Sistemas de Informação assume caráter de área científica.

Sistemas de Informação (SI) é uma área Sociotécnica que abarca, estruturalmente, três aspectos centrados em soluções computacionais, ou um ecossistema destas: comportamental ou humano; organizacional ou procedimental; técnico ou tecnológico. Diferente de Ciência da Computação (CC), que podemos dizer ser a irmã mais velha de SI, ela não necessita imprescindivelmente de elementos computacionais [Stair e Reynolds 2018]. Por exemplo, um comerciante, em uma feira de rua, que registra suas operações e transações comerciais em um caderno enquadra-se em SI, isto é, o aspecto técnico é o caderno, a caneta utilizada para escrever, o marcador de página, e suas especificidades técnicas. Partindo desde exemplo, há determinado consenso na literatura indicando que a adoção de soluções computacionais auxilia as operações e transações deste feirante [Stair e Reynolds 2018, Laudon e Laudon 2020]. Perceba que “adotar solução computacional” é uma prática, dotada de valores, e cabível de ponderação ética. Indicar que a adoção de solução computacional é “um avanço” ou “melhora” independente de contexto, i.e., variáveis espaciais e temporais; extrapola a essência ética, melhor adequada à uma normatividade tecno-determinística. Após uma ponderação ética desligada de pré-concepções, manter o uso do caderno ou optar por outra abordagem não-computacional são opções

¹O dicionário Caldas Aulete define o verbo engenheirar como “Exercer a profissão de engenheiro, exercer atividade relacionada a engenharia”. Disponível em: <https://www.aulete.com.br/engenheirar>. Disponível em: 01/09/2021

válidas e alinhadas com SI. Encaminhamentos análogos seriam, “por que ter um perfil em Rede Social Online (RSO)?”, “por que ter uma Alexa?”, “por que compartilhar determinada publicação em uma RSO?”, “por que programar em Python?”, ou “por que baixar um jogo pirata?”, direta ou indiretamente todos estas são reflexões éticas.

A análise pautada em SI precisa transcender, ou transbordar, o aspecto técnico. E apesar de cada um dos três aspectos ter mais ou menos importância dependendo da análise, busca-se uma compensação e parcimônia entre estes. O pensamento predominante e reducionista empurra o “fracasso” de uma solução computacional aos aspectos **não-técnicos**, i.e., os dois outros aspectos são os responsáveis pela “derrota” do sistema “perfeito”, supostamente a cultura, a educação, hábitos, costumes, leis, normas, processos, mentalidade, dentre outros. Analisando pelo viés de SI, a complexidade engloba também o técnico e busca uma visão ampla, e mesmo que, de fato, o ônus recai majoritariamente sobre os aspectos procedimentais ou comportamentais, o aspecto técnico não é “isento” ou livre de valores neste jogo de forças. Concluindo, enquanto a visão tecnicista positivista determinística empurra toda a valoração negativa aos aspectos não-técnicos, SI abraça uma ponderação plural e se afasta de, por exemplo, “é esse povo burro e mal-educado, essas leis ruins e esses processos mal estruturados que levaram nosso sistema incrível e perfeito ao fracasso”. Neste exemplo, pensando em SI, os elementos responsáveis pelo pensar-fazer deste sistema que falharam, ao desconsiderar ou menosprezar as especificidades de fato e de valor do aspecto comportamental (instrução e educação) e procedimental (leis e processos).

[Stair e Reynolds 2018] definem sistemas de informação como: “Conjunto de componentes inter-relacionados que coletam, processam, armazenam e disseminam dados e informações; um sistema de informações provê mecanismos para monitorar e controlar suas operações, garantindo que continue a cumprir seus fins e alcançar seus objetivos”² [tradução nossa]. Por exemplo, um automóvel com soluções computacionais embarcadas, instaladas ou embutidas; que ajudam o motorista a analisar o estado do próprio veículo, estacionar, acompanhar a previsão do tempo, interagir com mapas, evitar tráfego intenso, elaborar rotas e, em grau próximo ao máximo de automatização, dirigir. O pensar-fazer de todas estas funcionalidades é passível de ponderação ética, assim como suas aplicações, seus impactos e suas influências. Com este exemplo podemos ilustrar a diferença central nas imanências de CC e SI, enquanto CC pensa estas funcionalidades como fim em si, SI pensa este sistema enquanto aplicação, impacto e influência. Por exemplo, CC pensa em processamento gráfico para melhorar as interações com e representações de mapas; SI pensa sobre formas de comunicar estas informações gráficas acessíveis a determinado público-alvo, sobre as áreas marginalizadas ou periféricas ignoradas pelos mapeamentos automatizados ou exposição de dados relevantes para determinada necessidade.

Alguns outros sistemas de informação estão comumente presentes no nosso dia-a-dia, mesmo que o exemplo do automóvel recheie nossa imaginação com idealismo tecnológico fantasioso. Por exemplo, caixas de supermercado, farmácias, padarias, dentre outros utilizam uma categoria de sistema de informação chamada Sistema de Processa-

²“A set of interrelated components that collect, process, store, and disseminate data and information; an information system provides a feedback mechanism to monitor and control its operation to make sure it continues to meet its goals and objectives.” [Stair e Reynolds 2018]

mento de Transação (SPT). Até 2020 este era o exemplo mais simples e vinculado com nossa rotina diária, com a pandemia de COVID-19, e seus efeitos colaterais, a migração ao digital intensificou-se e o SPT mais usual, para muitos, se tornou sistemas de comércio eletrônico (*e-commerce*), e.g., Mercado Livre, Shopee, Amazon, Magazine Luiza, OLX, dentre outros.

Aspectos técnicos em SI não estão restritos ao software, compreendem também hardware, armazenamento de dados e redes [Stair e Reynolds 2018]. Por exemplo, um sistema de reconhecimento facial: (i) o software é, intuitivamente, o algoritmo de Inteligência Artificial (IA) que processa as imagens, os demais códigos envolvidos; (ii) hardware, as câmeras que capturam as faces; (iii) armazenamento de dados, tanto os dados armazenados após coleta pelas câmeras, quanto os dados presentes na base de dados dedicada à operação da IA; (iv) redes, os meios de transmissão de dados, consideremos uma estruturada cabeada neste cenário. Apenas este aspecto, cabe lembrar, não contempla o contexto de análise em SI, ainda faltam os aspectos procedimental e comportamental.

Neste sentido, precisamos nos afastar do padrão tradicionalista de sistemas de informação e pensar em SI para analisar casos por vias menos positivísticas ou determinísticas, como o fenômeno de compras online durante a pandemia de COVID-19. Pelo aspecto técnico isolado, implementar um sistema de informação *e-commerce* é a opção com melhores resultados e indicadores. Com este sistema todo o processo é automatizado, os clientes são cadastrados, o pagamento é efetivado com maior segurança, todas as transações são registradas, há possibilidade de controle de estoque, o cliente escolhe os produtos com mais precisão, e os dados estruturados neste *e-commerce* podem alimentar iniciativas de Inteligência do Negócio, servindo de insumo para melhorias do próprio sistema e no negócio [Stair e Reynolds 2018, Laudon e Laudon 2020]. Entretanto, diferente deste apogeu tecno-centrado, negócios se adaptaram de outras formas ³:

“Os empresários do setor de supermercados que não podem ou não querem investir nas vendas on-line porque sabem que elas não vão atingir seu público, tem usado de um recurso simples para oferecer um atendimento diferenciado, que são os pedidos pelo aplicativo de mensagens whatsapp. Muitos estabelecimentos passaram a oferecer o serviço no início da pandemia, pois sentiram que essa era uma necessidade para várias famílias e principalmente para os idosos que por medo de serem contaminados evitavam ao máximo sair de casa. [...] Sobre a possibilidade de migrar para as venda on-line nas plataformas especializadas, Adriane acredita que ainda não é o momento, pois teria que mexer muito na estrutura da loja. 'Nem para aplicativos de delivery de comida como o ifood. A maneira como está hoje é a ideal, nossa logística própria é mais barata e o cliente está satisfeito com o serviço', argumenta.”

Um sistema tradicionalmente utilizado para conversação foi expropriado de sua função primária e se tornou um “sistema de *e-commerce* informal”. Perceba que diversos dos benefícios, funcionalidades e operações que um sistema efetivo de *e-commerce* ofereceria, a opção de usar *Whatsapp* não oferece. Entretanto, para estas organizações esta opção foi um diferencial para lidar com o fenômeno do isolamento social e da quarentena,

³<https://opopularpr.com.br/compras-de-supermercado-pelo-whatsapp-cairam-no-gosto-popular/>. Disponível em: 01/09/2021.

esta é a abordagem em SI, de mãos dadas com os aspectos procedimentais, a organização adaptou seus processos e atendimento ao novo meio; e aspectos culturais, aproveitando a disseminação pervasiva já bem-estabelecida do aplicativo *Whatsapp*, os requisitos socio-culturais de seus clientes, e a facilidade de atendimento da plataforma, e.g., uma pessoa analfabeta pode realizar uma operação neste cenário, basta enviar um áudio detalhando seu pedido, o que seria impossível em um *e-commerce* tradicional.

São diversos os vieses éticos neste cenário: (i) do negócio em utilizar o *Whatsapp* como veículo de compras; (ii) dos clientes de respeitar a quarentena e o isolamento social; (iii) dos funcionários do negócio, menos expostos a entidades externas e evitando possíveis aglomerações nas lojas; (iv) da empresa *Facebook*, em perceber que sua plataforma *Whatsapp* está sendo utilizada como operação de varejo e permitir que assim seja; (v) dos entregadores, que se colocam em risco para realizar as entregas; (vi) do governo, que perdeu a oportunidade de auxiliar ativa e instrutivamente os negócios à Transformação Digital; (vii) dos próprios serviços de *e-commerce*, que perderam espaço para outros aplicativos aleatórios, ao invés de implementar módulos de integração entre sistemas (sistemas de sistemas). Todas estas foram vertentes de ação cabíveis de ponderação de valores. Todavia, nem todas seguiram análises racionais e fundamentadas, pré-requisito básico para Ética [Fieser 2020, Ferraz 2014, Gibson 2014].

Neste trabalho consideraremos SI enquanto área, não analisaremos sistemas de informação especificamente, como tradicionalmente tidos. Nosso escopo primário é Ética, aqui indissociável de SI. Ademais, dos cinco cursos oficiais que compõe oficialmente a área de Computação no Brasil ⁴, SI apresenta uma interdisciplinaridade e heterogeneidade mais adequado ao currículo suficiente e necessário para um diálogo com Ética. Mais do que isso, refletir sobre valores e não-neutralidade em SI, do pensar e do fazer.

Seguindo esta linha de raciocínio, na Seção 6.2.3 analisaremos, com propriedade, porque a episteme clássica e tradicional de CC é prejudicial à Ética Computacional, ou deficitária à Ética de forma geral.

Nosso objetivo nesta obra é propor alguns apontamentos sobre o atravessamento entre Ética e SI, na concentração do pensar-fazer. Isto é, está fora do nosso escopo tratar de Ética na epistemologia de SI. Por exemplo, não analisaremos os valores práticos da tripartite de aspectos proposto por [Stair e Reynolds 2018]. De forma iterativa, e interessante, encontramos diversos sistemas de informação que auxiliam na pesquisa em SI, e podemos analisá-los pelo viés de SI também, assim sucessivamente. Por exemplo, detectores de plágio, poucos são os detectores de plágio em Língua Portuguesa, prejudicando o uso por interessados sem determinado nível de letramento em Inglês, idioma dominante. Em certos casos os detectores de plágio online têm suas interfaces traduzidas automaticamente, gerando resultados repletos de erros na interface em Língua Portuguesa.

Esta obra está estruturada da seguinte forma: a Seção 6.2 traz as fundamentações teóricas mais específicas sobre Éticas e as áreas de aplicação aqui cobertas; na Seção 6.3 cobriremos aspectos éticos relacionados primariamente ao fazer; na Seção 6.4 cobriremos aspectos éticos relacionados primariamente ao pensar; a Seção 6.5 lida com alguns

⁴Ciência da Computação (CC), Engenharia da Computação (EC), Engenharia de Software (ES), Licenciatura em Computação (LC) e Sistemas de Informação (SI). [Brasil 2016a]

extremos e casos polêmicos para criticar senso comum; e a Seção 6.6 apresenta a discussão final e conclusão.

6.2. Fundamentação Teórica em Ética

A Ética tem seus primórdios seminais na Grécia, a partir de pensadores como Sócrates, Platão e Aristóteles [Ferraz 2014]. Esta é uma visão hegemônica, por uma perspectiva de colonização, com foco na trajetória europeia e estadunidense. Extrapolando estas fronteiras, outros pensadores de locais diversos analisaram os valores associados às ações humanas, podemos citar na China, Confúcio e Lao Tzu [Gibson 2014]. Em África esta rastreabilidade se torna mais complexa [Gibson 2014, Gyekye 2010], algumas das razões são (i) a forte tradição da oralidade e transmissão de conhecimento por meios sem registro longínquo, e.g., escrita; (ii) resquícios de conflitos excessivos no continente que acarretaram em prejuízos socioculturais, e.g., guerras internas ou colonização radical; (iii) Ética, tal qual dialogamos, é um construto oriundo dos gregos, e diversos povos de África pensavam “Ética” de forma própria. Seria absurdo ou pejorativamente discriminatório conjecturar que os povos de África, continente plural, vasto e composto de diversos países, povos e nações, não debruçou-se à reflexão de seus próprios costumes, hábitos e tradições, com a possibilidade de que a vastidão de pensadores e estudiosos africanos partiram de pontos iniciais diversos e diferentes dos europeus/gregos. Por exemplo, um dos elementos africanos associados à Ética é *Ubuntu* [Gibson 2014]. [Gyekye 2010] é uma base inicial para adentrar o universo da Ética e Moralidade africana.

Este parágrafo introdutório teve o objetivo de esclarecer que parâmetros e indicadores associados à Ética variarão geográfica e espacialmente. Acostumados à Ética proveniente de raízes europeias e estadunidenses, operar e lidar com Filosofia Moral em espaços predominadas por culturas externas a estas culminará na necessidade de apreender novas vertentes éticas.

Neste trabalho abordaremos a vertente dominante no contexto brasileiro, oriundo da Grécia e disseminado e aceito pelas fundações europeias e estadunidenses. Por exemplo, ao viajar para a China, há probabilidade de que não apenas as normas e ditames morais sejam diferentes, assim como as visões de ética em si. Com a globalização, velocidade e disponibilidade de acesso pelo mundo, você pode operar, e.g., trabalhar; em organizações em quase todo planeta Terra e alocado no Brasil. Outro exemplo, a Estônia oferece a possibilidade de residência eletrônica (*e-Residency*)⁵, e do Brasil você pode realizar diversas operações e procedimentos na Estônia, e na União Europeia. Neste caso concreto, é necessário pensar nas premissas éticas e morais estonianas de antemão.

Dada esta delimitação contextual, seguimos às fundamentações Éticas pertinentes.

6.2.1. Ética

Diversas são as definições de Ética, detalharemos e dialogaremos com a seguinte: “Seguimento da filosofia que se dedica à análise das razões que ocasionam, alteram ou orientam a maneira de agir do ser humano, geralmente tendo em conta seus valores morais.”⁶

⁵e-resident.gov.ee/start-your-eu-company-with-e-residency/. Disponível em: 01/09/2021

⁶<https://www.dicio.com.br/etica/>. Disponível em 01/09/2021

As fundamentações em Ética serão a partir de [Ferraz 2014, Gibson 2014, Fieser 2020, Deigh 2010] A Ética, também conhecida como Filosofia Moral, analisa as ações humanas e seus valores. Nos estenderemos na Ética propriamente dita, generalista, pois a mesma serve de alicerce e é aplicável às demais sub-categorias de Ética Aplicada que iremos tratar.

Um dos pensamentos seminais que seguem da Grécia à atualidade é associado a Aristóteles, a tripartite das faculdades humanas. Estas faculdades são composta por três níveis hierárquicos, **desejo, emoção e razão**. Nossas ações são orientadas por estes níveis, um diferencial entre os animais e seres humanos é a orientação e acesso ao nível mais elevado, onde animais estão restritos aos níveis de desejo e emoção. Por este mesmo motivo deixamos de atribuir valores às ações de animais, assim como responsabilização. Por exemplo, se um roedor mastiga a fibra ótica, o mesmo o fez estimulado por desejo ou emoção, não razão, sendo assim não há juízo de valor, i.e., ele não está errado, é mau ou injusto. Mesmo que essa ação desestabilize toda a Internet de uma cidade, ele tanto não refletiu eticamente sobre ela, assim como não é dotado de capacidade mental de fazê-lo posteriormente.

De forma breve e resumida, o nível do desejo envolve instintos e sensações; o nível da emoção envolve as subjetividade e sentimentos; o nível da razão envolve o que os filósofos consideravam o ápice da mente humana, tornando-se sábios. “O sujeito que age de forma moralmente correta, por seu turno, é aquele que permite que sua parte superior domine, isto é, sua razão” [Ferraz 2014], esta citação expõe como a linha grega predominante pensava na relação entre a Ética e a razão. Assim como coberto pela definição acima, a análise ética é a análise das razões, i.e., ela está acima das emoções e dos desejos. Atente-se que são “as razões”, pois as pessoas arrazoam suas ações por razões tipificadas diferentemente [Deigh 2010].

Trazendo ao contexto de SI, quando interagimos com uma desinformação em uma RSO, podemos ser compelidos, por desejo ou emoção à compartilhá-la, entretanto a Filosofia Moral recomenda que a interação informacional seja regida pela razão. E mesmo que esta informação esteja em conformidade com nossa visão de mundo, apelativa aos nossos instintos, ou afinadas aos nossos sentidos, compartilhar esta informação de forma irracional é uma ação moralmente irresponsável. Por exemplo, compartilhar apenas pela chamada.

O sujeito da Ética é a Moralidade, composta pela Moral. Enquanto a moral é subjetiva e inerente ao indivíduo, a ética é objetiva e remete à coletividade. Cada indivíduo tem suas ações guiadas pelos seus valores morais no geral, como a definição aponta. A Ética analisa, a Moral age. Esta distinção é importante, pois a perturbação ética parte da moral, por exemplo, conforme o processamento de dados pessoais intensificou-se e começou a gerar impactos institucionais negativos graves, o debate ético rumou a institucionalmente considerar diversas práticas imorais. Este fenômeno associado à moralidade sobre o processamento de dados pessoais que motivou o surgimento da Lei Geral de Proteção de Dados (LGPD) [Brasil 2018a].

Nós refletimos sobre nossas ações a partir da ética. A ação, consciente ou inconscientemente, é guiada pela nossa moral primariamente. A Ética consiste da moderação entre a nossa moral, i.e., nossa subjetividade de valores; e a moral de terceiros, nor-

malmente coletiva. A combinação desta moderação compartilhada entre os indivíduos compõe a moralidade. Conceitualmente, “antiético” não é uma qualidade, ao considerarmos que a ética é uma análise das ações, i.e., “não ética” (aética) é a ação realizada sem reflexão racional. Informalmente, podemos qualificar ações deliberadas e sistematizadas pela moralidade como imorais como antiéticas, e o mesmo para ações éticas. Por exemplo, um casal troca fotos de nudez em aplicativo enquanto namorados, e após o término uma das partes resolve deliberadamente vazar as fotos da outra parte na Internet. Neste caso, podemos considerar essa ação como antiética, pois a moralidade bem-difundida e estabelecida considera a mesma negativa. Isto é, dentro da flexibilidade de opções viáveis neste cenário, não há quase nenhum espaço para esta condução de valores.

Por exemplo, o caso dos termos de uso, políticas de privacidade e compartilhamento de dados do *Whatsapp*^{7 8}. Sobre o aceite dos termos:

“Aceitando os novos termos, o usuário autoriza que o Facebook absorva informações das pessoas que utilizam o aplicativo para impulsionar a monetização de conteúdos nas redes sociais - ou seja, gerando anúncios mais personalizados, grande parte da renda da Facebook Inc, empresa que comanda as duas redes sociais.”⁷

“Essa atualização da política de privacidade não chega a ser uma mudança de fato (nas regras de compartilhamento de dados). O que muda, na verdade, agora é que as pessoas perceberam que elas não tinham o direito de se opor ao compartilhamento dos seus dados enquanto elas estavam usando os serviços de WhatsApp’, afirma Rielli.”⁸

Perceba que esta é uma ação deliberada, analisada e avaliada pela empresa. Não há necessidade ao funcionamento ou obrigatoriedade técnica para este compartilhamento de dados entre dois serviços computacionais diferentes. Isto é, o *Facebook* não depende dos dados do *Whatsapp* para operar, e vice-versa. E mesmo o argumento de que “melhora a experiência do usuário através de uma interoperabilidade de dados e melhoria em predições, perfilamento e recomendações” não se sustenta, porque é ausente ao usuário a opção de não compartilhar, i.e., ele é levado, por padrão, à esta opção porque o não-compartilhamento é opção fora do padrão ou não disponível. Sendo assim, considerando que os usuários são titulares de seus dados, que a ação não é necessária aos requisitos primários das aplicações, que a LGPD indica que esta linha de ação é inadequada e que o usuário não tem poder sobre como são utilizados seus próprios dados, podemos pensar nesta prática com valores morais alinhados à antiética.

Dando prosseguimento, a Ética consiste de uma negociação dialógica entre as subjetividades morais do indivíduo, internamente; e das objetividades dos valores morais da realidade externos a ele. Este é um ponto filosoficamente demasiado complexo para

⁷<https://www.opovo.com.br/noticias/tecnologia/2021/06/01/apos-polemicas-whatsapp-desiste-de-limitar-aplicativo-para-quem-nao-aceitou-novos-termos.html>. Disponível em: 01/09/2021

⁸<https://www.uol.com.br/tilt/noticias/bbc/2021/01/13/novas-regras-para-o-whatsapp-4-perguntas-sobre-as-mudancas-no-aplicativo.htm>. Disponível em: 01/09/2021

ser detalhado aqui, entretanto entender a relação entre si e o meio é uma base essencial à Ética.

Por exemplo, você mora em um condomínio e, justificando necessidade de mais segurança, vai instalar uma “câmera inteligente” que vai sensorear e capturar as movimentações fora do seu apartamento, logo em cima da sua porta principal. As gravações serão armazenadas na nuvem e você terá acesso no seu computador e celular. Dito isso, seus vizinhos de porta concordam com a ideia de serem filmados? As pessoas que passam ou vão ao seu apartamento, concordam? O entregador do *iFood* tem o direito de solicitar que você apague os registros da aparição dele das suas gravações? Se a polícia requisitar essas gravações para investigar seu vizinho, você cede ou tem o dever de ceder? Se a gravação capturou seu filho quebrando uma janela do condomínio, aí ele mente dizendo que não foi ele e o próprio condomínio inicia uma investigação, você cede as gravações? Ou a integridade moral do seu filho está acima disso? Note que nenhuma dessas questões é trivial de refletir sobre, apesar de serem todas associadas à ética e que você deveria ter pensado sobre elas antes de comprar e instalar esse hardware.

Que tal pensarmos em um exemplo menos fantasioso? Uma Alexa ⁹. Imagine ir na casa de um amigo e ter uma Alexa gravando o que vocês falam, ou um funcionário da Amazon escutando a conversa de vocês? Pior, com a possibilidade deste sensoreamento de áudio ser ativado “do nada”? Há alguns anos atrás essa opção seria considerada absurda, hoje muitos compram uma Alexa tendo consciência de que estarão jogando uma roleta russa com a sua privacidade. E não apenas com a sua, também com a de terceiros. Moralmente, os valores associados ao fenômeno de bem-estar enquanto “vigiado” ou “capturado” implicitamente está mudando. Eticamente, ainda nos cabe refletir sobre os valores morais associados com esta mudança de comportamento e possível paradigma.

Duas características da ponderação ética são necessárias aqui. A auto-preservação e a intenção própria. Argumentativamente, há uma necessidade **racional** de deliberarmos sobre individualismo e coletivismo. Ética não se trata apenas do que nós queremos, muito menos deve ser alinhada à nossa visão de mundo ou realidade; ao mesmo tempo a ética não deve ser um artifício de auto-depreciação ou coerção oculta ou latente. O indivíduo deve refletir sobre seu papel na comunidade coletivizada na qual participa, assim como a associação de suas predileções individuais neste contexto.

Dois extremos, “vou piratear este software porque se eu posso ‘não pagar’ por nada sempre, vou escolher esta opção” e “piratear software é desrespeitar toda a equipe de desenvolvimento, prejudicar toda uma indústria, dar um exemplo ruim para os outros e essa opção nunca deve ser sequer imaginada, por ninguém”. Se você não tiver uma percepção ampla e heterogênea do dilema ético, e.g., pirataria; que está tratando, sua análise será comprometida por um viés e parcialidade restritos. Todas as nossas escolhas, e ações oriundas delas, são enviesadas e parciais, e isso faz parte da imanência em ser humano; o ideal, a partir da Ética, é de que haja uma minimização de ruído e de viés através de uma visão panorâmica e holística.

Um exemplo clássico externo à Computação, não é **necessário** fazer uma análise

⁹<https://www1.folha.uol.com.br/tec/2019/04/empregados-da-amazon-ouvem-o-que-voce-diz-para-alexa.shtml>. Disponível em: 01/09/2021

imparcial e neutra sobre o nazismo e “apontar o que o regime nazista fez de bom, correto e justo” para deliberar eticamente sobre o nazismo, o regime nazista, os nazistas ou os impactos e influência do nazismo. Apenas de repetir essa proposta em voz alta e refletir alguns segundos sobre ela já se percebe sua irrazoabilidade. Por este mesmo motivo que elementos de Memória Social, como estátuas, nomes de ruas, motivos culturais, símbolos, dentre outros, são proibidos de associação com elementos nazistas. Por exemplo, o caso de jogos onde uniformes de oficiais do regime nazista e suásticas são banidos [Garon 2018].

Finalmente, há a auto-preservação interseccional identitária enquanto pessoa inserida em uma comunidade ou sociedade desigual. Por exemplo, quando se trata de racismo algorítmico [Silva 2020] ou misoginia e machismo na comunidade de Computação [Ribeiro et al. 2020]. Nossas características étnicas, raciais ou de gênero fazem parte de quem somos, e negar isso é auto-anulação de um elemento de nossa subjetivação. Você pode ser motivado pelas ameaças à sustentabilidade social, inclusive a si mesmo, e por combater as desigualdade, seja em Computação, e respectiva comunidade, ou através da Computação, utilizando-a como artifício para esta empreitada. Entretanto, esta linha de abordagem ética precisa ser trilhada com cautela, porque podemos acabar guiados principalmente pela emoção ou desejo... e o pensar-fazer formal deve ser majoritariamente **guiado** pela razão, sem necessariamente desconsiderar desejos ou emoções.

Concluindo a fundamentação em Ética, este domínio apresenta suas próprias bases de fundamentação e princípios. A coleção de possibilidades para dialogar com Ética é vasta. A divisão tradicional separa a Ética em Metaética, Ética Normativa, Ética Aplicada e, mais recente, Ética Descritiva. Focaremos na Ética Aplicada, neste caso na Ética Aplicada à Computação e SI, caso necessário e oportuno, recorreremos às demais. Por exemplo, refletir eticamente sobre as especificidades e diferenciações entre a LGPD, no Brasil, e a *General Data Protection Regulation* (GDPR), na União Europeia (UE), é objeto da Ética Descritiva. Outros construtos aplicados inerentes à Ética serão cobertos na Seção 6.2.3, dedicada ao escopo principal desta obra.

6.2.2. Ética Computacional

[Johnson 2008] define Ética Computacional como: “O estudo das questões éticas que surgem como consequência do desenvolvimento e implantação de computadores e tecnologias de computação. Envolve duas atividades. Uma é identificar e colocar em foco as questões e problemas que se enquadram em seu escopo, a consciência da dimensão ética de uma situação particular. A segunda é fornecer uma abordagem para essas questões, um meio de aumentar nossa compreensão e encontrar maneiras de chegar a soluções sábias para esses problemas.”

As fundamentações em Ética Computacional serão a partir de [Barger 2008, Johnson 2008, Masiero 2013, Spinello 2020, Baase e Henry 2017] e materiais ou cadernos de estudos sintetizados sobre o tema do *Massachusetts Institute of Technology* (MIT)¹⁰. A Ética Computacional, um dos ramos da Ética Aplicada, analisa as ações humanas e seus valores aplicadas do domínio da Computação. Fazendo uma analogia com Programação Orientada a Objetos, entre classes, Ética é uma super-classe, Ética Aplicada é uma sub-

¹⁰https://www.cs.uct.ac.za/mit_notes/ethics/. Disponível em: 01/09/2021

classe da Ética, e a Ética Computacional é uma sub-classe da Ética Aplicada. Isto é, os níveis mais baixos herdam as características daqueles acima. Por exemplo, Privacidade é tanto um tema recorrente em Ética, quanto em Ética Aplicada, quanto em Ética Computacional. Um exemplo de tópico inerente à Ética Aplicada é a privacidade na relação profissional entre clientes e prestadores de serviços, a Ética Médica (outra sub-classe da Ética Aplicada) refletirá sobre dilemas diferentes da Ética Computacional.

Há, também, o debate sobre a Ética Computacional apresentar especificidades e propriedade enquanto sub-classe da Ética Aplicada, i.e., “O domínio de Computação tem dilemas especializados tais quais suas especificidades sejam passíveis de uma ramificação própria da Ética Aplicada? Ou são dilemas genéricos e não cabem ramificação específica?”. Neste trabalho dialogamos com as percepções de [Johnson 2008] e com [Moor 1985], sendo este último um dos artigos seminais ao tema, e consideramos que a Ética Computacional tem relevância tal qual apresente especificidades e propriedades enquanto sub-classe da Ética Aplicada.

Sem prolongar de forma extensa, há uma incoerência na ideia de Ética Computacional enquanto restrita a juízo de fato, principalmente porque a Computação é uma atividade, também, social. Diferentes de fenômenos da natureza, por exemplo, artefatos e pensamentos computacionais não são “naturais”. Isto é, se formos analisar Segurança em Tecnologia da Informação e Comunicação (TIC), devemos pensar em como a comunidade/sociedade dialoga com o conceito de segurança para, por exemplo, determinar parâmetros e critérios. Normalmente análises tratam dos cibercriminosos como “elementos etéreos”, e.g., o *e-commerce* da Renner foi alvo de um *ransomware*¹¹, este ataque não “veio do nada”, isto é, para analisar este caso eticamente precisamos tratar de aspectos alheios aos técnicos. Caso a análise fique limitada aos aspectos técnicos, deixa de ser uma análise ética e recai em irrealidade, porque o conjunto de normas e os humanos envolvidos são elementos da realidade concretos envolvidos no caso.

É neste sentido que posicionamos SI, do leque de cursos oficiais em Computação [Brasil 2016a], como adequada para contemplar a Ética Computacional. E, neste sentido, é difícil conceber a ideia de que uma formação com o currículo de CC, com pouquíssima ou nenhuma base de aspectos alheios ao técnico, instruirá ou conscientizará um especialista de forma a habilitá-lo na pluralidade e heterogeneidade que a premissa Ética pressupõe. O que não significa que hajam problemas éticos inerentes ao aspecto técnico, principalmente ao que tange juízos de fato, entretanto a possibilidade de viés ao pensamento de que “tudo se resume à técnica e o resto é secundário ou descartável” é maior, considerada a esmagadora maioria de episteme técnica.

No quesito da Ética e neste caso, cabe diferenciar juízo de valor e juízo de fato [Lacey 2006], de forma simples (i.e., deixaremos o debate kantiano de fora, por exemplo). O juízo de fato engloba a objetividade e o juízo de valor, subjetividade. Neste sentido, aspectos técnicos isolados são governados por juízo de fato. Por exemplo, a complexidade ou tempo de um algoritmo; taxa de transmissão de uma rede; capacidade de armazenamento de um disco rígido; ou quantidade de núcleos de um microprocessador. Isolados, estes elementos parecem “deslocados” de todos os elementos ou atributos cabíveis de val-

¹¹<https://www1.folha.uol.com.br/mercado/2021/08/renner-fica-mais-de-24-horas-com-sistema-fora-do-ar-apos-ciberataque.shtml>. Disponível em: 01/09/2021

oração moral. Entretanto, a pessoa, coletivo ou organização que pensou-fez estes itens são dotadas de valores morais; as consequências de impacto, influência ou aplicação destes itens envolvem valores morais, já que envolvem indivíduos; em certo nível consciencial, estes itens reforçam valores morais [Donia e Shaw 2021, Pereira et al. 2015]. Nenhum destes itens é elaborado ou desenvolvido para si, i.e., como fim; e só pensando nesta premissa já percebemos que a análise ética necessita de aspectos diversos. Sem estes aspectos diversos ela está isolada ao juízo de fato, isolada ao juízo de fato ela aliena o item de seu contexto, alienado do contexto não há Ética.

Considerando temas gerais de Ética Computacional, temos: (i) Privacidade, atualmente Proteção de Dados incluso; (ii) Cibercrimes ou Segurança; (iii) *cyberbullying*; (iv) Propriedade Intelectual; (v) Acesso Indevido; (vi) Aspectos Trabalhistas; (vii) Códigos de Ética ou Conduta. São alguns temas específicos por autores:

- [Barger 2008]: Internacionalismo; Robótica; Poder. Propõe um tópico chamado Computação Parasita (ou Parasítica) que diz ser “literalmente” inerente à Computação, i.e., não é generalizável para demais sub-classes de Ética Aplicada.
- [Johnson 2008]: Aspectos societários; Vigilância; Normatividade Digital (Ordem Digital).
- [Spinello 2020]: Poder; Liberdade de Expressão.
- [Baase e Henry 2017]: Erros, falhas e riscos; Liberdade de Expressão; Normatividade Digital (Avaliando e Controlando Tecnologia).
- [Masiero 2013]: Internet. O autor recorre a diversos casos, obras e tópicos de outros autores.

Nestas obras supracitadas não há distinção de alocação entre áreas acadêmico-científicas distintas, e.g., Códigos de Ética e Conduta são oriundos da ES. O escopo adquire uma complexidade “frankensteiniana” e transdisciplinar, onde as fronteiras de onde começa a termina Ética Computacional são nebulosas e ambíguas. Por exemplo, [Masiero 2013] delimita a Ética Computacional ao excluir tópicos onde computadores apareçam de forma utilitária. Entretanto, na mesma obra, [Masiero 2013] instrumentaliza casos computacionais de forma utilitária. Por exemplo, o caso onde uma professora vê um aluno assistindo conteúdo pornográfico no computador.

Igualmente, nestas obras supracitadas, as análises práticas são baseadas em casos, em sua maioria reais. E estes casos não estão limitados à forma utilitária, ou limitados ao aspecto técnico, ou apenas a juízo de fato; eles estão relacionados com aspectos comportamentais e procedimentais. Se pensarmos no próprio tópico de Computação Parasita, de [Barger 2008], há transbordamento do aspecto técnico.

Este debate das fronteiras da Ética Computacional é bastante extenso para endereçarmos aqui, todavia há uma série de questões que emergem desta linha. O encaminhamento que conjecturamos adequado é da inter/transdisciplinaridade, i.e., para englobar a Ética Computacional com propriedade precisamos dialogar com outras áreas do conhecimento que não apenas a Computação pura.

Por outro lado, há o juízo restrito, ou próximo de ser considerado restrito, do aspecto técnico isolado. Neste ponto podemos considerar os juízos de valor como secundários ou inadequados. Por exemplo, um código de software para funcionalidade X construído com linguagem de programação Y pode ser operacional e funcional. Entretanto, ele pode ter problemas técnicos, onde neste cenário podemos também analisar os valores morais de seu desenvolvedor. O mesmo ainda é um código ruim, apesar de não estar “errado”. Esta ação de construir este código é passível de análise ética, distante de um juízo de valor e com maior proximidade de juízo de fato. Se este mesmo código de software fosse construído com linguagem de programação Z, seria factualmente bem melhor para a funcionalidade X, o juízo sobre **este artefato computacional** é de fato. Paralelamente, podemos analisar o porquê da escolha do desenvolvedor pela linguagem Y, e não pela Z, e daí extrair um juízo de valor. Por exemplo, (i) “escolhi Y e não Z porque gosto mais da Y”, agiu por afeição e não por razão; “escolhi Y e não Z porque não sabia da existência de Z”, agiu por ingenuidade e ignorância do amplo aspecto técnico. Cabe lembrar, códigos de software não “desenvolvem a si mesmos” do nada.

Há uma precariedade de trabalhos formalmente, especificamente e explicitamente dedicados à Ética Computacional no Brasil, produzidos por brasileiros ou em português brasileiro [Carvalho et al. 2021b]. É uma das características, que prejudica este cenário, é de que diversas obras dedicadas à Ética Computacional são rotuladas como “Tecnologia e Sociedade”, “Computação e Sociedade” ou “Aspectos Sociais da Computação” e similares. Um exemplo é [Ramalho 2013], com uma obra intitulada “Tecnologia e Sociedade” inteiramente dedicada à Ética Computacional. Apesar de não ser um demérito ou prejuízo de qualidade da obra, desvirtua-a de seu tópico primário. E, como já mencionamos aqui, uma análise da melhor qualidade plausível em Ética Computacional deve abarcar tópicos como “Sociedade” e “Aspectos Sociais”, sem necessidade de uma significação terminológica ambígua.

Neste sentido, as obras supracitadas tratam da análise de casos externos ao Brasil, em grande maioria estadunidenses. [Masiero 2013], também, é fortemente ancorado em casos e apontamentos estadunidenses. Sobre o tema Ética Computacional, é limitada a quantidade de obras analisando casos brasileiros, assim como obras em português brasileiro produzidas por brasileiros.

6.2.3. Ética em Sistemas de Informação

Se já é desafiador e complexo encontrar fronteiras para Ética Computacional, i.e., onde a área começa e onde ela termina e o que ela engloba, com propriedade; esse cenário é intensificado ao refletirmos sobre Ética em SI. Como dito por [Bock et al. 2021]: “Ainda não há concepção ampla, coerente e aceita sobre como o campo entende a sua agenda ética, e o que deve ser coberto por esta agenda” [tradução nossa]¹².

Nesta Seção nos baseamos em quatro Revisões Sistemáticas da Literatura (RSL) dedicada especificamente em Ética e SI: [Bock et al. 2021, Paradice et al. 2018, Stahl 2012, Mingers e Walsham 2010]. Estes trabalhos cobrem o tema com abrangência e oferecem uma abordagem bola de neve para trabalhos relacionados.

¹²“There is as yet no coherent and widely accepted conception of how the field understands its ethical agenda, and what it takes to be covered by that agenda”. [Bock et al. 2021]

Além da ausência de concepção ampla, coerente e aceita, a partir destas quatro obras fomos incapazes de (i) encontrar um conceito consensual, sendo um aspecto negativo exclusivamente associado à apropriação categórica, sem ser um revés da argumentação ou conteúdo em si; (ii) perceber as diferenças entre Ética em SI e Ética Computacional. [Bock et al. 2021], em sua tentativa de diferenciação e categorização, anuncia que nenhum outro sub-domínio da Ética é devoto ao tópico principal da Ética em SI, que seria “desenvolvimento, uso e manutenção de sistemas de informação em/para ação social em sistemas organizacionais”¹³.

Considerando a definição de sistema de informação como exposta na Introdução e esta tentativa de restrição de escopo por [Bock et al. 2021], permanecemos incapazes de diferenciar Ética em SI de Ética Computacional, propriamente. Em sentido amplo, nos parece que Ética em SI é uma sub-classe ou irmã de Ética Computacional dedicada ao contexto organizacional, o que pode ser explicado pela recorrência de Ética Empresarial (*Business Ethics*). Esta, por sua vez, é uma sub-classe da Ética Aplicada que trata dos padrões organizacionais contemporâneos, seus princípios, conjuntos de valores e normas que governam as ações e comportamentos de indivíduos, ou grupos, em um ambiente organizacional formal [Moriarty 2021]. Considerada esta linha de raciocínio, a Ética em SI é uma combinação indissociável destas duas sub-classes.

Tendo a sua base e fundação ancorada nos Estados Unidos da América, país com uma forte tradição de organizações privadas, percebemos socioculturalmente disseminada a ideia de “Ética Empresarial” baseada em Empresa, definida como “Sociedade organizada para a exploração de indústria ou comércio; com a finalidade de obter um rendimento monetário através da produção de bens ou de serviços”¹⁴. Entretanto, a Administração Pública¹⁵ é constituída de organizações formais relacionando-se dialogicamente com sistemas de informação, assim como empresas. Definir uma “organização formal”, por outro lado, é uma tarefa ainda mais complexa, e.g., um grupo estruturado em uma RSO, uma comunidade em um fórum online ou uma guilda em um jogo digital online, dentre outros. Se nos mantivermos restritos à ideia associada ao conceito de empresa, limitaremos o escopo de Ética em SI e correremos o risco, pensando pelo viés do Interacionismo Simbólico, de negligenciar ou ocultar o envolvimento de demais tipos de “organizações formais” perceptíveis. Principalmente Organizações Públicas, vinculadas à Administração Direta.

A partir de um viés crítico, convoca-se à uma maior importância e conscientização ética no domínio de SI e indica-se que este debate está aquém do esperado ou desejado. [Stahl 2012] anuncia que a comunidade de SI não atingiu seu potencial em relação ao tema e há uma carência de trabalhos em níveis mais elevados de abstração. O debate deve ultrapassar meras preferências morais, e deve ocorrer em conjunto com teorias éticas subjacentes. [Paradice et al. 2018] indica que mais casos de ensino e instrução são necessários, assim como mais cursos ou disciplina que lidam especificamente com dilemas de Ética Organizacional em contextos tecnológicos. Encoraja a elaboração de mais casos e cursos. [Bock et al. 2021] é, de longe, o mais crítico e incisivo nos resulta-

¹³“[...] development, use, and maintenance of information systems in and for the social action systems of organizations.”

¹⁴<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/empresa>. Disponível em: 01/09/2021

¹⁵pt.wikipedia.org/wiki/Administração_pública. Disponível em: 01/09/2021

dos de seu artigo, onde indica que há relativamente pouca comunicação científica sobre Ética em Pesquisa em Sistemas de Informação (PSI); não há conceito unificado de Ética em SI; há pouco trabalho tanto sobre design de artefatos gerado por PSI quanto trabalhos empíricos. Isto é, há espaço e chamada de avanço no tema, e isso considerando que estes autores **não** cobriram trabalhos brasileiros ou o contexto brasileiro, onde há ainda menos casos registrados e estudados, cursos e comunicações científicas dedicadas.

O atravessamento entre o domínio da Ética e domínio de SI (ou Computação) requer um diálogo entre os conceitos, construtos e princípios de ambos. Para este fim, consideram-se mais adequadas três abordagens, (i) Deontologia, (ii) Teleologia ou Consequencialismo ou (iii) Ética das Virtudes ou Comunitarismo. Outras frentes fogem destas vertentes clássicas, como Floridi e a Ética da Informação [Tavani 2002]. Propõe atribuir ou transferir valor moral à solução computacional, de forma que a mesma é interpretada com devido protagonismo, ou até caráter de ator, e não como mera “ferramenta”.

Vamos ser breves nas explicações destas três abordagens, que são tanto detalhadas nas obras de referência desta Seção, 6.2.3, como na Seção de Ética Computacional, 6.2.2.

Consequencialismo, e Teleologia, sustenta que o prejuízo/dano geral deve ser minimizado e o bem geral deve ser maximizado pelas ações corretas. Podemos pensar que são as consequências das ações que contam ou que as ações são julgadas com base no grau de bondade/bem que resultam. Esta linha é seguida por utilitarianistas, que podem ser baseados em duas vertentes: baseado em ação, julga as ações de indivíduos ou grupos; baseado em regras, julga os resultados da adoção de determinado conjunto de regras ao bem geral ou bem maior. Sendo uma linha simples de apreender as premissas iniciais, tem suas contrapartidas. No domínio da Computação, é bastante complexo prever ou antecipar as consequência de ações, principalmente em longo-prazo. Como o caso citado por [Spinello 2020], sobre a abertura da plataforma para desenvolvedores externos no contexto de jogos:

“Por alguns anos, a plataforma de desenvolvedor do Facebook hospedou vários jogos populares, incluindo Farmville e Candy Crush. Os clientes do Facebook concordaram em dar a esses desenvolvedores de jogos acesso aos seus dados em troca de jogar esses jogos. Porém, não havia proteções para o reaproveitamento desses dados coletados pelos desenvolvedores. Os algoritmos extraíam itens como mensagens e fotos dos usuários. Um desenvolvedor de jogos usou dados do Facebook para construir perfis não autorizados de crianças em seu próprio site. O Facebook permitiu o compartilhamento de seus dados de clientes sem um sistema para evitar abusos.”¹⁶.

Deontologia, ao invés de tratar da consequência da ação, o foco está na ação em si. Enquanto o consequencialismo permite a ideia de “os fins justificam os meios”, a deontologia repudia categoricamente este pensamento. Esta linha é idealista e baseada fortemente nas ideias do filósofo Immanuel Kant, que propôs o Imperativo Categórico. São

¹⁶“For a few years the Facebook developer platform hosted several popular games including Farmville and Candy Crush. Facebook customers agreed to give these game developers access to their data in exchange for playing these games. However, there were no protections for the reuse of these data collected by the developers. Algorithms were extracting items such as users’ messages and photographs. One game developer used Facebook data to construct unauthorized profiles of children on its own website. Facebook had allowed for the sharing of its customer data without a system to prevent any abuses.” [Spinello 2020]

duas diretrizes, a máxima de que devemos ultrapassar a preocupação conosco e basear nossas ações aos outros; a máxima deve ser universal e aplicável a todos. Por exemplo, se considerarmos que mentir é errado, nunca devemos mentir, independente da circunstância. Isto é, consideremos um cenário de uma ditadura, se você for preso pelas forças ditatoriais e questionarem onde estão seus companheiros, você deverá falar a verdade, pois esta é a máxima moral.

Uma segunda vertente deontológica é baseada na ideia de contratos sociais. Estabelecemos contratos sociais com diversas entidades, entre elas o Estado. Neste sentido, coletivamente a sociedade mantém e protege os direitos dos indivíduos e os indivíduos protegem este contrato social, e as ações são julgadas em termos de manutenção e reforço destes direitos. Isto é, a premissa está baseada nos ditames deste contrato social e as consequências são secundárias, enquanto as determinações deste contrato social são a máxima. Nesta vertente, o mesmo se estende para termos de uso e políticas de privacidade em sistemas de informação, ao concordar com este contrato são estabelecidos os parâmetros morais.

Ética das Virtudes, ou Comunitarismo, trata de virtudes individuais ou coletivas. Nesta vertente o foco desvia da ação em si e desloca-se ao indivíduo ou coletivo específico. Deve ser alcançada uma elevação, consistindo do balanço entre extremos. Por exemplo, há um meio-termo entre sempre mentir e sempre falar a verdade e o ideal é alcançá-lo. A frônese (*phrónesis*) que nos permite balancear as virtudes, sendo a sabedoria prática ligada à dialética das ações. Aristóteles formulou a Ética das Virtudes, sendo um empirista associou o alcançar a média áurea, este meio-termo, com a experiência e prática humana. Por exemplo, a Coragem está entre a imprudência e a covardia, devemos buscar sempre agir nesta intermediação, todavia em determinados momentos é possível que precisemos apelar a um ou a outro.

Alcançar a média áurea através da frônese, para Aristóteles, constitui no objetivo de vida a ser alcançado e mantido para o que chama de “uma vida boa e virtuosa”. MacIntyre estendeu a proposta para uma abrangência comunitária, o foco da média áurea é individual e o comunitarismo é baseado em uma média áurea comunitária. Pela filosofia do Comunitarismo, as comunidades desenvolvem seus códigos e julgamentos éticos coletivamente através de socialização e consequente desenvolvimento. Comunidades diferentes geram suas próprias práticas e padrões e é impossível ultrapassar todas as tradições para um ponto de vista eterno e universal. Isto é, contextos específicos atrelados a comunidades específicas que são sempre relativos à “uma vida boa e virtuosa”. A crítica principal à Ética das Virtudes e ao Comunitarismo é sua tendência à relativização. Em uma configuração multicultural e globalizada, a Ética Relativista pode oferecer desafios excessivamente complexos e perigosos.

Um exemplo de choque de valores morais ocorreu entre a plataforma *Facebook* e o Ministério da Cultura (MC) brasileiro, em um caso de 2014¹⁷. O MC publicou uma fotografia de dois indígenas do grupo Botocudo em homenagem ao Dia do Índio, em sua página institucional na plataforma. Esta fotografia, exposta na Figura 6.1, mostra uma indígena de seios despídos. A plataforma excluiu a imagem argumentando que feria suas

¹⁷<https://canaltech.com.br/redes-sociais/ministerio-da-cultura-acusa-facebook-de-censura-por-foto-bloqueada-39633/>. Disponível em: 01/09/2021

políticas internas de publicações dos usuários. O próprio ministro da cultura entrou em contato com a plataforma e emitiu um comunicado: “tenta impor ao Brasil, e às demais nações do mundo onde a empresa opera, seus próprios padrões morais agindo de forma ilegal e arbitrária [...] fere a Constituição da República; o Marco Civil da Internet; o Estatuto do Índio e a Convenção da Unesco sobre Proteção e Promoção da Diversidade e das Expressões Culturais. Também desrespeita a cultura, a história e a dignidade do povo brasileiro”. Posteriormente o *Facebook* reverteu a exclusão e bloqueio do conteúdo, entretanto foi causado um mal-estar. Apesar da discussão sobre a validade da decisão, se estava em conformidade com as regras da plataforma ou não, demonstrou como comunidades diferentes estabelecem e fazem valer valores morais diferentes.



Figure 6.1. Indígenas da etnia Botocudo, por Walter Garbe [Garbe 1909]

Considerados os elementos conceituais e de fundamentação sobre Ética e SI, seguimos para as bases do pensar, a Ética em Pesquisa.

6.2.4. Ética em Pesquisa

O ser humano desenvolve pesquisas em todas as áreas do conhecimento e como as questões éticas associadas às novas tecnologias extrapolam as fronteiras da tecnologia e atingem a sociedade de maneira direta, existem fatores que estão diretamente relacionados às questões humanas. [Padilha et al. 2005] ressaltam que os pesquisadores estão atentos às questões burocráticas da pesquisa, como os termos, autorizações e aspectos operacionais para a realização da pesquisa, e muitas vezes, não dão a devida atenção ao seu elemento principal: o ser humano.

É preciso garantir alguns aspectos relacionados diretamente com os participantes da pesquisa, sejam os pesquisadores, ou as pessoas que são o foco da intervenção de pesquisa. É importante garantir que nenhum desses indivíduos seja exposto a riscos desnecessários e nem que sofram algum tipo de discriminação. Para que isso seja contemplado no projeto é essencial que sejam adotadas metodologias, materiais e equipamentos adequados.

Diferentes métodos e metodologias de pesquisas são descritas na literatura e nem sempre consideram participações diretas de pessoas ou animais. [Cesário et al. 2020] consideram que a pesquisa científica opera em dois níveis, teórico e empírico. O nível teórico desenvolve teorias abstratas sobre um fenômeno social ou natural e o nível empírico está voltado a realizar teste dos conceitos teóricos para definir as observações da realidade. Os autores destacam que as abordagens das pesquisas podem ser quantitativas ou qualitativas. Em geral, as pesquisas quantitativas envolvem procedimentos experimentais, com resultados numéricos passíveis de comprovação estatística. Já as qualitativas consideram a bibliografia de uma área, estudos que coletam opiniões pessoais, observações com participação direta do pesquisador ou não, pesquisas etnográficas, dentre outras. Estas pesquisas podem ser analisadas de forma indutiva, entretanto, podem ter resultados transformados em números para uma análise mais representativa dos resultados.

Considerando as técnicas que envolvem a participação de pessoas, destacamos alguns modelos de pesquisas [de Pádua 2019]:

- Pesquisa experimental que avaliam os resultados da aplicação de procedimentos inéditos, ou não, em geral, de forma controlada, em que acontecem situações e ações baseadas em hipóteses bem fundamentadas. As pesquisas na área da Saúde exploram bastante este modelo;
- Pesquisa de levantamento de dados, que estuda o objeto da pesquisa em uma população e usa fortemente questionários fechados, ou entrevistas baseadas em roteiros. Este modelo é bastante utilizado nas pesquisas de áreas sociais e humanas. Mas, também é muito utilizado em avaliações de qualidade de software;
- Pesquisa com estudo de casos, que avaliam resultados de poucos objetos de pesquisa, considerando casos isolados ou pequenos grupos. Esta abordagem pode ser utilizada em muitas áreas e explora estudos com uma pessoa, ou grupo específico.

O pesquisador deve escolher a melhor estratégia para o fenômeno que quer estudar e que possa alterá-la de forma a atingir os objetivos de forma precisa. Os métodos adotados seguem etapas bem definidas, associadas à área da pesquisa que, em geral, consideram regras e modelos replicáveis.

Esse rigor científico vai gerar a confiança nos resultados obtidos, que serão passíveis de análises detalhadas, considerando pesquisas correlatas como fonte de comparação. E essa confiança está relacionada com o cuidado no levantamento, armazenamento e tratamento dos dados para que gere resultados relevantes para a sociedade.

Esse conjunto de práticas bem fundamentadas determinam resultados confiáveis,

que respeitam diretrizes éticas, que respeitam a vulnerabilidade dos participantes e respeitam suas liberdades, garantindo que danos previsíveis serão evitados.

A Universidade Federal do Rio Grande do Sul (UFRGS) lançou um guia para integridade de pesquisas científicas, que descreve detalhes de vários aspectos contemplados sobre este tópico ¹⁸.

6.3. Ética fazendo Sistemas de Informação

Esta Seção é dedicada ao “fazer” SI, em amplo espectro. Fazer SI ultrapassa o simples desenvolver ou construir código, como conduz o senso comum. De início, fazer SI envolve também hardware, armazenamento de dados e redes, assim como software; em seguida, demais elementos não-primários em relação com TIC e fundamentais ao fazer são considerados, como documentação (incluindo comentários em código), arquitetura empresarial, qualidade(s), relações inter-pessoais, versionamento, interação com o cliente ou usuário, aspectos políticos, precificação, dentre outros. Todos estes elementos estão fora do escopo de TIC propriamente dito, entretanto são essenciais ao fazer um artefato de TIC.

[Stair e Reynolds 2018] dividem SI em sentido amplo e em SIBC (Sistemas de Informação Baseado em Computador). Um SIBC obrigatoriamente terá TIC computacional, um SI não necessariamente. Por exemplo, é comum encontrar consultórios médicos ou de dentista brasileiros com os cadastros e dados pessoais dos clientes em papel, e.g., fichas pautadas; e estes **também** são sistemas de informação, por exemplo, estão sob auspícios da LGPD mesmo sendo meio físico. Nesta obra focamos em SIBC, com especificidades e complexidades específicas inerentes à Computação. Um SIBC contempla elementos computacionais de TIC (aspecto técnico) + pessoas (aspecto comportamental) + procedimentos (aspecto procedimental).

6.3.1. Códigos de Ética

Ética está relacionada com justiça social e vai além da lei, é o nosso compromisso com uma sociedade melhor. Um ambiente ético é criado efetivamente pelos valores cultivados pelas pessoas, prevalecendo atitudes com ênfase na honestidade, na palavra de cada pessoa, e no seu exemplo, com comportamentos de igualdade, justiça e confiança. No entanto, com o passar do tempo, podem ser criadas leis para inibição de comportamentos considerados prejudiciais à coletividade, e estabelecem-se penalidades para estes comportamentos, e.g., roubos ou assassinatos.

A *Association for Computing Machinery* (ACM) lista alguns princípios éticos gerais, moralmente intuitivos para muitos de nós ¹⁹: contribuir para o bem-estar humano e da sociedade, sabendo que todas as pessoas são partes interessadas à Computação; evitar danos; ser honesto e digno de confiança; ser justo e agir de forma a não discriminar; respeitar o trabalho requerido para produzir novas ideias, invenções, obras criativas e artefatos computacionais; respeitar a privacidade; honrar a confidencialidade. Mesmo sendo tão genéricos e aplicados à vida cotidiana, podemos estabelecer relações entre estas diretrizes e nosso comportamento enquanto profissionais e usuários de sistemas de

¹⁸<https://www.ufrgs.br/propesq1/propesq/wp-content/uploads/2020/06/guia-para-integridade-em-pesquisa-cientifica.pdf>. Disponível em: 01/09/2021

¹⁹<https://www.acm.org/code-of-ethics>. Disponível em: 01/09/2021

informação.

Além dos princípios gerais que norteiam o bom funcionamento social, existe também a ética de determinados grupos ou locais específicos. Por exemplo, ética médica, ética profissional, ética empresarial, ética educacional, ética nos esportes, ética jornalística, ética na política, etc. Pensando na área de SI, como já mencionamos, não existe ainda uma “Ética” consolidada, observamos uma relação dialógica entre a Ética em Computação e os chamados códigos de ética, ou de conduta.

[Masiero 2013] explica que o profissional em Computação é quem projeta e desenvolve artefatos computacionais, produzindo os resultados de um processo de desenvolvimento de software: documentos de especificação de requisitos, modelos de análise e projeto, planos de teste, código fontes, manuais, planos de manutenção, dentre outros. Podemos acrescentar aqueles que trabalham com infraestrutura, redes e hardware de computadores em geral. Este profissional é normalmente qualificado através de cursos superiores de bacharelado em Computação, além dos cursos técnicos em Informática.

Sabemos então o que faz um profissional de Computação, mas esta profissão ainda não é regulamentada oficialmente no Brasil, e isso significa entre outras coisas que não há um código de ética formal para guiar a conduta destes profissionais. Código de Ética Profissional é um conjunto de normas éticas, que devem ser seguidas pelos profissionais no exercício de seu trabalho.

A maioria das profissões têm o seu próprio código de ética profissional derivadas da ética geral, e muitas vezes incorporado à lei pública. Nesses casos, os princípios éticos passam a ter força de lei. Perceba que, mesmo nos casos em que esses códigos não estão incorporados à lei, seu estudo tem alta probabilidade de exercer influência, por exemplo, em julgamentos nos quais se discutam fatos relativos à conduta profissional. Ademais, o seu não cumprimento pode resultar em sanções executadas pela sociedade profissional, como suspensão temporária ou definitiva do direito de exercer a profissão. Estes códigos são elaborados pelos respectivos Conselhos que representam e fiscalizam o exercício de cada profissão. Por exemplo, o Conselho Regional de Engenharia e Agronomia (CREA), e o Conselho Regional de Medicina (CRM). Então não existe um código de ética da Computação, com força de repressão institucional, pela ausência de um conselho regulamentando esta profissão.

No entanto, algumas organizações ligadas à área de Computação no mundo, tais como a ACM, o *Institute of Electric and Electronic Engineers* (IEEE-CS), desenvolveram projetos para criação de códigos de ética unificados. As recomendações oferecidas por estes códigos vão desde valores relativos ao interesse individual até importância para a sociedade em geral.

O código de ética da ACM¹⁹ é mais geral e abrange também profissionais que atuam em hardware, infraestrutura e redes de computadores. Já o código de ética da IEEE-CS/ACM²⁰ foi proposto em conjunto pelas duas entidades, e é voltado para a Engenharia de Software, e portanto, para profissionais que desenvolvem software.

[Anderson et al. 1993] também discutem em sua publicação o uso prático do

²⁰<https://www.computer.org/web/education/code-of-ethics>. Disponível em: 01/09/2021

código da ACM. Considere visitar estes sites e ler atentamente os princípios, ou imperativos, definidos. Os principais personagens, a quem os códigos de ética se referem, são: profissionais de computação, sociedade em geral, usuários de sistemas, empregadores, empregados, clientes, colegas de trabalho e organizações. Uma das recomendações existente no código da ACM é: “Avaliar de forma abrangente e profunda os sistemas de computação e seus impactos, incluindo análise de riscos potenciais” [tradução nossa].

O *Computer Ethics Institute* criou os famosos 10 mandamentos da Ética Computacional (“*The Ten Commandments of Computer Ethics*”²¹). Esta lista resume as questões mais sensíveis que abrangem a Ética Computacional:

1. Não usarás um computador para causar mal a outrem.
2. Não interferirás no trabalho de computação de outrem.
3. Não bisbilhotarás os arquivos de computação de outrem.
4. Não usarás um computador para roubar.
5. Não usarás um computador para dar falso testemunho.
6. Não copiarás e nem usarás software proprietário pelo qual não tenhas pago.
7. Não usarás os recursos computacionais de outrem sem a devida autorização ou compensação.
8. Não te apropriarás do produto intelectual de outrem.
9. Considerarás as consequências sociais do programa que estás escrevendo ou do sistema que estás projetando.
10. Usarás sempre um computador por caminhos que garantam a consideração e o respeito pelos teus semelhantes.

6.3.2. Propriedade Intelectual

Propriedade Intelectual (PI), que compreende Direitos Autorais e Patentes, é um tema amplamente presente na literatura de Ética Computacional, arriscaríamos apontar sua presença em todas as obras do tema [Barger 2008, Johnson 2008, Masiero 2013, Spinello 2020, Baase e Henry 2017]. Sua importância acentuada se estende à SI, diretamente presente em [Stahl 2012]. PI, quando aplicada à Computação, é tópico-afim de SI, enquanto aspecto procedimental primariamente.

Aqui nesta Seção seguiremos uma abordagem crítica em relação à PI associada à SI. Em todas estas obras supracitadas a argumentação é favorável à PI, i.e., estão carregadas de valores morais enviesados e pré-concebidos. Sendo assim, nos parece que a essência da Ética é deturpada para uma normatividade moral conduzida. Se Ética trata de dilemas e PI é moralmente o ponto final e não há contra-argumentação ou viés contrário, então deixa de ser um dilema ético e se trata absoluto moral. Ser crítico, contrário ou

²¹<http://computerethicsinstitute.org/publications/tencommandments.html>. Disponível em: 01/09/2021

complementar a certas especificidades de PI é diferente de ser “absolutamente contrário” à PI. Ao contrário, as literaturas supracitadas expõe linhas argumentativas absolutamente favoráveis à PI e avessas a fenômenos conflitantes com este tópico.

O fenômeno popular que infringe a moralidade de PI é conhecido como “pirataria”, formalmente rotulado de contrafação²². Plágio e contrafação são diferentes, neste sentido separamos os dois em categorias normalmente já separadas, contrafação em fazer e plágio em pensar. Em sentido amplo, ambos podem se manifestar independente da categoria de prática.

Considerando o foco contextual, países e blocos políticos específicos possuem legislações e moralidades diferentes ao tratar de objetos morais e ética. As obras supracitadas são fortemente ancoradas no cenário estadunidense, i.e., em determinado nível impossíveis de generalização ou transferência ao cenário brasileiro. Neste caso, algumas produções brasileiras se destacam no enfoque e profundidade do assunto, apesar de descoladas do tema da Ética Computacional. Recomendaremos alguns trabalhos abrangentes e posicionados no Brasil.

Em questão institucional, o Instituto Nacional da Propriedade Intelectual²³ apresenta uma pluralidade de materiais e guias sobre o tema, atualizados. Capítulos de livro, [Nunes e Pinheiro-Machado 2017], “Propriedade Intelectual e Busca de Informação Tecnológica na área da Computação” e [Kon et al. 2020], “Direitos autorais, licenças e patentes”. Em formatos diversos, o Almanaque para popularização de ciência da computação sobre Propriedade Intelectual²⁴, como ilustrado na Figura 6.2. O website Um Caso de Arte²⁵, do Instituto de Tecnologia & Sociedade do Rio, com material do especialista em Direito Autoral Sérgio Branco.

Por outro lado, podemos também citar produções críticas a determinados aspectos ou à totalidade da situação atual da PI. As principais críticas tratam de desigualdades de privilégio e poder com ênfase em aspectos econômicos. [Marques 2016] aponta um caso concreto, da empresa Ultron na década de 1980, e critica a propriedade intelectual sob vieses colonizadores:

“Através dos direitos à propriedade intelectual, o primado da origem é mais ou menos sub-repticiamente evocado e traduzido para garantir o primado do centro sobre a periferia, da Europa sobre a América Latina, do primeiro sobre o terceiro mundo ou do ‘colonizador’ sobre o ‘colonizado’.” [Marques 2016]

Seguindo na linha de raciocínio de [Marques 2016], [Darch 2004] trata de como o fenômeno de Desigualdade Digital (*Digital Divide*) se relaciona com os países do Sul Global, como Brasil, apontando efeitos negativos da PI que os EUA e países da Europa impõe em sua respectiva agenda sobre o tema. Uma das saídas apontadas é o Software Livre, situando SI no olho do furacão. Enquanto este trabalho data de 2004, em 2021 ele

²²<https://jus.com.br/artigos/55180/notas-sobre-o-plagio-e-a-contrafacao/2>. Disponível em 01/09/2021

²³<https://www.gov.br/inpi/pt-br>. Disponível em: 01/09/2021

²⁴<http://almanaguesdacomputacao.com.br/serie3baixa.html>. Disponível em: 01/09/2021

²⁵<https://umcasodearte.com.br>. Disponível em: 01/09/2021



Figure 6.2. Recorte. Série 3: Propriedade Intelectual. Volume 11: Propriedade Intelectual no Comércio Eletrônico ²⁴

ainda é um tema em disputa ²⁶, tratando da indústria da produção acadêmica diretamente relacionada à PI e Ciência Aberta.

[Marques 2014] traz as teorias da Economia Política da Informação e do Conhecimento e o pensamento de Karl Marx para a discussão do papel da informação e do conhecimento na atualidade para propor um olhar crítico à PI. Defende a necessidade de debater as contradições do tema, pois os atores que participam da disputa discursiva enviesadamente e acriticamente defendem PI, financiados com recursos públicos, entre eles Governos e instituições de pesquisa e ensino. [Perelman 2003] também acompanha esta linha de raciocínio, indicando que a PI proporciona um mundo de litígios excessivos, violações intrusivas de privacidade, a destruição do sistema de ensino superior, interferência na pesquisa científica e uma distribuição desigual de renda.

Para concluir, um dos casos mais polêmicos envolvendo PI ocorreu na UE, tomando proporções transnacionais midiáticas em 2017. Em 2013 a UE encomendou uma ampla pesquisa, financiada com dinheiro público, sobre Direito Autoral (*Copyright*) e consequências de contrafação e pirataria ²⁷. As questões que nortearam o estudo foram: (i)

²⁶<https://www.timeshighereducation.com/opinion/push-open-access-making-science-less-inclusive>. Disponível em: 01/09/2021

²⁷<https://edri.org/our-work/did-the-eu-commission-hide-a-study/>. Disponível em: 01/09/2021

Como as violações de direitos autorais online afetam as vendas de conteúdo protegido por direitos autorais? (ii) Quanto os infratores de direitos autorais online estão dispostos a pagar por conteúdo protegido por direitos autorais? ²⁸. Em resumo, o estudo conclui que a pirataria e contrafação são positivas à indústria da música, livros e jogos digitais; sendo negativa apenas à indústria de filmes, principalmente *blockbusters*. Fugindo do senso comum e da agenda moralista, é um discurso fora da curva para reflexão.

Neste sentido, o debate ético consiste da análise de lentes diversas sobre tópicos em disputa, evitando os extremos. Estes trabalhos críticos não necessariamente repudiam integralmente a base ou os conceitos de PI, sim trazem outros olhares, por perspectivas alheias à deontologia categórica, consequencialismo punitivo e ao idealismo determinístico tradicionais deste debate.

6.3.3. Leis e normas

A relação dialógica entre a moralidade e as tecnologias é complexa, de forma que nos ausentaremos desta profundidade. Brevemente, as tecnologias interagem com a sociedade ou coletividade e estas respondem às tecnologias, assim sucessivamente. [Moor 1985] debate a relação interativa entre a massa crítica deliberando criticamente em proporções cada vez maiores ou mais intensas e como este fenômeno afeta a moralidade, i.e., os valores morais estabelecidos e acordados entre as partes interessadas. Nesta Seção focaremos em normas jurídicas e técnicas.

De início, leis e normas têm relação direta com deliberação ética, explícita ou implícita, entretanto não é garantia de que sejam, efetivamente, morais. O exemplo clássico é relacionado à escravidão, em determinado momento histórico a escravidão foi legal, isto não significa que a escravidão é moral. Atualmente, escravidão é um dilema ético superado sem possibilidade de debate enviesado, enquanto outros dilemas atuais são relacionados com aborto, liberação animal, aquecimento global, desinformação, dentre outros. Isto é, seguir leis e normas não configura atitude ética, e agir eticamente não significa cegamente e literalmente seguir todas as leis e normas.

Neste sentido, determinadas leis podem ser negligenciadas via de regra, por diversas partes envolvidas, popularmente dizemos que a lei “não pegou”. Como na lei de São Paulo ²⁹ que proíbe o uso de celulares em agências bancárias, e mesmo que pouquíssimos respeitem, os próprios sistemas bancários incentivam que os clientes usem celulares nas suas agências, para interoperação entre as aplicações do celular com os terminais eletrônicos, e.g., leitura de *token* para habilitar as operações de Pix no celular, como o cliente vai fazê-lo sem usar o celular na agência? Por que há esta funcionalidade implementada se a lei impede? Considerando que a lei serve apenas aos indivíduos, podemos considerar o banco conivente nestes casos?

Coletivamente, estabelecemos acordos morais entre nós. Por exemplo, quando as pessoas dividem e compartilham regras e critérios em casa, como respeitar os idosos, não falar palavrão, almoçar em determinado horário, limpar o próprio quarto, dentre outros. Em determinado momento podemos definir contrapartidas caso os acordos sejam des-

²⁸https://netzpolitik.org/wp-upload/2017/09/displacement_study.pdf. Disponível em: 01/09/2021

²⁹<http://g1.globo.com/bom-dia-brasil/noticia/2012/10/uso-de-celular-nos-bancos-e-proibido-mas-clientes-nao-respeitam-lei.html>. Disponível em: 01/09/2021

cumpridos, como ficar sem mesada, não comer sobremesa. E podemos definir “melhores práticas” para determinadas ações, por exemplo, a melhor forma de limpar as janelas, uma forma ideal de fazer o arroz, a melhor coleira para passear com o cachorro, como limpar a chupeta do bebê, dentre outros. A norma jurídica, i.e., lei, estabelece elementos de coerção para moldar o comportamento dos sujeitos através de imperativos de conduta, como se espera e deseja em uma célula da coletividade, e transgredir uma lei culmina em uma sanção [Bobbio 2019, Dimoulis 2016], é uma regra mandatária e punitiva. A norma técnica é “um documento, produzido por um órgão oficial acreditado para tal, que estabelece regras, diretrizes, ou características acerca de um material, produto, processo ou serviço. A obediência a uma norma técnica, tal como norma ISO ou ABNT, quando não referendada por uma norma jurídica, não é obrigatória”³⁰.

Há uma diferença de reação entre uma lei (norma jurídica) e norma técnica. A transgressão de uma lei culmina em punição ou efeito negativo, a adequação e conformidade com normas técnicas culmina em potenciais recompensas ou efeitos positivos. Podemos perceber que a década de 2010 foi um período de intensificação da relação dialógica entre a computação, e seus impactos ou influências, com a tecnologia computacional no Brasil. Uma das leis resultantes é a Lei Geral de Proteção de Dados (LGPD) [Brasil 2018b]. Em relação a juízos de valor e juízos de fato, normas técnicas são fortemente associadas a juízos de fato; enquanto leis, juízos de valor. Um exemplo de norma técnica associada à LGPD é a família de normas da ISO/IEC 27000³¹, com foco em Sistema de Gestão de Segurança da Informação, sendo a ISO 27001 passível de certificação regulamentada e as demais recomendações de boas práticas. Determinadas relações inter-organizacionais formais dependem da conformidade com normas técnicas, e certificações são formas bem-estabelecidas de expor que se seguem as diretrizes determinadas pela norma.

Existem diversas interpretações e críticas sobre a relação entre as leis e a sociedade. Em sua maioria, a literatura relacionada à Ética Computacional é parcial favorável a leis e normatizações mandatárias [Johnson 2008, Baase e Henry 2017]. Pensando pelo viés da Teoria do Conflito e com uma visão crítica, [Mascaro 2013] associa a normatização jurídica, tanto formulação quanto estrutura, ao estado burguês e manutenção do sistema capitalista. No Estado burguês, as leis são pensadas para privilegiar os detentores e favorecidos pelo Capital, e manter a estrutura dominante para uma minoria. Por exemplo, em versão inicial a LGPD definia que decisões automatizadas poderiam ser revisadas por revisão humana caso solicitado, algum tempo depois houve uma alteração crucial, as decisões automatizadas seriam revisadas de forma também automatizada. A decisão foi criticada por especialistas³², considerada um retrocesso ao avanço esperado pela lei:

“Enfim, considerando a ausência de transparência e os efeitos perversos das análises exclusivamente automatizadas, a revisão humana acerca das decisões tomadas a partir de dados pessoais é uma salvaguarda constitucional e democrática que tenderia a

³⁰https://pt.wikipedia.org/wiki/Norma_técnica. Disponível em: 01/09/2021

³¹https://pt.wikipedia.org/wiki/ISO/_IEC_27000. Disponível em: 01/09/2021

³²<https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-lgpd>. Disponível em: 01/09/2021

evitar danos aos cidadãos mais carentes que não possuem outro acesso a bens que geram cidadania e ficam dependentes da análise do “sistema”.

Na era da tecnologia, é preciso realçar a importância do direito à explicação e à revisão humana das decisões automatizadas que impactam a vida das pessoas.

Desta feita, tendo em vista o atual estado da arte, o veto presidencial se revela um grave equívoco de entendimento político acerca da importância da obrigatoriedade da revisão humana. Pelo menos até que as IAs evoluam e possam demonstrar de forma segura as motivações de suas conclusões, sujeitando-se a algum tipo de controle.”³²

Neste sentido, refletimos sobre a relação e influência de agentes que operam em favor do Capital, como bancos e agentes do mercado financeiro, que poderiam ter perdas ou riscos financeiros substanciais com a obrigatoriedade da revisão humana. Uma preocupação das empresas em relação à conformidade, adequação e aderência às leis é pela gestão de sua imagem institucional [Bioni 2019], e neste momento de ascensão do debate sobre racismo algorítmico [Silva 2020] é demasiado arriscado ser rotulado como uma empresa opressora e racista através de suas operações automatizadas, e envolver atores humanos nestes casos culmina na possibilidade de erros comunicativos ou envolvimento de subjetividades humanas, impossíveis considerando atores computacionais.

Podemos esperar uma profusão de normas jurídicas relacionadas à Computação, e sua aplicação, na década de 2020. Já no início temos a “Lei das Fake News” (PL 2630/2020)³³; ou a PL 21/2020, que “Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências”³⁴. Aprovadas ou não, demonstram a preocupação associada à moralidade relacionada com estes dois fenômenos da computação, seja desinformação desenfreada ou inteligências artificiais opacas ou nocivas à sustentabilidade social.

Assim como a Lei Sarbanes-Oxley (SOX)³⁵ influenciou o cenário computacional transnacional na década de 2020 [Mingers e Walsham 2010], a LGPD está influenciando o período atual desta obra. Tão abrangente e pervasiva, a LGPD é um marco para o processamento de dados e proteção dos mesmos. Se associarmos a LGPD [Brasil 2018b] com a definição de sistema de informação [Stair e Reynolds 2018], observamos que ambas são indissociáveis, tanto no pensar quanto no fazer.

Nos ausentamos de uma profundidade analítica técnica sobre a LGPD, considerando que há uma abundância de obras dedicadas ao tema. Entretanto, há uma carência de obras técnicas se comparada com a quantidade de obras contextuais e conceituais. Sobre a LGPD, de forma ampla, recomendamos as obras e cursos do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio)³⁶ e Data Privacy BR³⁷. São encontrados cursos e obras gratuitos e pagos, com certificado.

³³<https://www25.senado.leg.br/web/atividade/materias/-/materia/141944>. Disponível em: 01/09/2021

³⁴<https://www.camara.leg.br/propostas-legislativas/2236340>. Disponível em: 01/09/2021

³⁵https://pt.wikipedia.org/wiki/Lei_Sarbanes-Oxley. Disponível em: 01/09/2021

³⁶<https://itsrio.org/pt/home/>. Disponível em: 01/09/2021

³⁷<https://dataprivacy.com.br>. Disponível em: 01/09/2021

6.3.4. Tomada de Decisão Ética em SI

[Moor 1985] discute que um problema típico na Ética Computacional surge porque existe uma lacuna política sobre como a tecnologia computacional deve ser usada. A computação nos oferece novos recursos e estes, por sua vez, nos dão novas opções de ação. Políticas de conduta para diversas destas situações ou as políticas existentes são inexistentes ou inadequadas. Desta forma, uma tarefa central da Ética Computacional é determinar o que devemos fazer nesses casos, i.e., formular políticas para orientar nossas ações. Uma outra dificuldade é que, juntamente com a política, existe uma lacuna conceitual. Embora um problema na Ética Computacional possa parecer inicialmente claro, uma pequena reflexão pode revelar um conflito conceitual. O que é necessário em tais casos é uma análise que forneça uma estrutura conceitual coerente dentro da qual formular uma política de ação [Moor 1985].

Desde o ano 2000, [Bynum 2000] já faz uma reflexão sobre o futuro do desenvolvimento tecnológico. No futuro, como resultado da invenção de máquinas “imitadoras da vida”, haverá mudanças dramáticas nas políticas e práticas sociais, na lei, na economia e nas relações humanas. Considere, por exemplo, algumas das implicações econômicas dos autômatos computadorizados. O desenvolvimento de tais máquinas “dá à raça humana uma nova e mais eficaz coleção de escravos mecânicos para realizar seu trabalho”. Mas o trabalho escravo efetivamente elimina muitos empregos humanos, tornando as pessoas sem valor no mercado de trabalho. Mesmo o trabalho intelectual de um cérebro humano não está imune à substituição por máquinas. Se construirmos máquinas “cujas capacidades intelectuais duplicam as dos seres humanos”, quais serão as consequências sociais e éticas? Será que tais máquinas têm um propósito próprio que rivaliza com o dos humanos? Eles terão o direito de “florescer” e ter “princípios de justiça” como os humanos? Será que máquinas cujas capacidades intelectuais excederem as dos seres humanos terão direitos que, portanto, sobrepoem-se aos direitos humanos? Estas são apenas algumas das questões éticas óbvias que surgiriam da existência de tais autômatos [Bynum 2018].

Percebemos, por estes exemplos, questões em SI cabíveis de atenção. É complexo proceder diante de determinadas polêmicas ou discussões, então para ajudar no raciocínio sobre estes assuntos, [Laudon e Laudon 2020] sugerem um passo-a-passo para deliberar sobre questões éticas em Computação que pode ser bastante útil na prática. São eles:

1. Identifique e descreva claramente os fatos: descubra quem fez o quê, quando e como; a visão correta da situação ajudará a encontrar uma solução.
2. Defina os conflitos ou dilema e identifique os valores envolvidos: uma questão ética envolve direções de ação opostas que envolvem valores importantes (por exemplo, liberdade, privacidade, proteção de propriedade, etc.)
3. Identifique os interessados: observe quem são os personagens interessados no desenlace da questão ética.
4. Identifique alternativas razoáveis a adotar: liste opções possíveis para o desfecho da questão (que nem sempre vão satisfazer a todos os interessados).

5. Identifique potenciais consequências das suas opções: questione-se o que ocorreria ao longo do tempo com cada uma das alternativas listadas.

A sua deliberação sobre uma questão ética é completa ao utilizar todos os elementos coletados e produzidos neste passo a passo para, então, adotar sua posição. Neste momento, talvez você precise tomar como base algum(ns) princípio(s) filosófico(s). No âmbito filosófico, os princípios, enquanto regidos pelas leis morais, são valores que o indivíduo considera adotar de acordo com o que diz sua consciência. Alguns exemplos são: Idealismo, Realismo, Pragmatismo.

Para o idealista, a bondade é encontrada na perfeição, no imaterial ou ideia de alguma coisa; e o mal é a distorção do ideal. Portanto, imperativos éticos nunca podem mudar. Idealistas julgam a ação em si e não suas consequências. Como estabelecer uma fundamentação moderna para a ética que seja, simultaneamente, atemporal e universal (válida sempre em todos os lugares)? Se uma ação não é correta para todos, então não é correta para ninguém. Se todos fizessem isso, a organização ou sociedade poderiam sobreviver? Immanuel Kant é um exemplo célebre de idealista.

Para o realista, todo conhecimento é obtido através dos sentidos. A bondade é encontrada ao se viver uma vida de virtude em harmonia com a natureza. O homem em estado natural é governado pelas leis da natureza onde todos são iguais e nenhum indivíduo deve prejudicar o outro. Tem o direito de usar a força para defender os seus direitos, mas não para ferir os direitos dos outros indivíduos. Seu direito termina quando começa o do outro. John Locke é um exemplo célebre de realista.

Para o pragmático, a realidade é um processo, mudança. Acontecimento, ou seja, experiência. O valor de algo é determinado unicamente em termos de sua utilidade para alcançar algum fim. Portanto, o fim justifica os meios. Princípio da Utilidade: a ação dotada de “maior valor ético” é aquela que maximiza a felicidade geral e minimiza a dor; Cálculo que visa o maior benefício para o maior número de pessoas (maximização do prazer). Realize a ação que produza a melhor relação custo/benefício. John Stuart Mill é um exemplo célebre de pragmatista.

Indo um pouco mais além, para [Laudon e Laudon 2020], “a introdução de nova tecnologia de informação tem o efeito de ondas concêntricas que suscitam novas questões éticas, sociais e políticas, as quais precisam ser tratadas nos níveis individual, social e político”. Eles explicam que questões éticas estão conectadas a questões políticas e sociais e se propaga nos níveis individual, da sociedade e do estado.

[Laudon e Laudon 2020] ressaltam que cinco dimensões morais nos questionam sobre a presença dos Sistemas de Informação e da TI na sociedade atual. (1) Direitos e obrigações sobre a informação: que direitos sobre a informação relativa a si próprios os indivíduos e organizações possuem?; (2) Direitos e obrigações sobre a propriedade: como os tradicionais direitos de propriedade intelectual serão protegidos em uma sociedade digital?; (3) Prestação de contas e controle: quem deverá prestar contas e ser responsabilizado por danos causados aos direitos individuais e coletivos sobre a informação e propriedade?; (4) Qualidade do sistema: que padrões de qualidade de dados e sistemas devem ser exigidos para proteger os direitos individuais e a segurança da sociedade?; (5) Qualidade de vida: que valores devem ser preservados em uma sociedade baseada na informação e con-

hecimento? Quais instituições devem ser protegidas contra a violação? Que valores e práticas culturais são apoiados pela nova tecnologia de informação?

6.4. Ética pensando Sistemas de Informação

Esta Seção é dedicada ao “pensar” SI, em amplo espectro. Desenvolvemos sistemas de informação simultaneamente enquanto pensamos, aqui lidamos com o pensamento consciente, englobando estudos, ensaios, entre outros. Pesquisar SI consiste em analisar a realidade envolvendo os três aspectos conceituais contidos na definição da disciplina.

Enquanto área, são encontradas inconsistências e incoerências entre cursos de universidades, cidades, regiões e países diferentes [Kohun et al. 2012], a “identidade” que SI assume também está associada à entidade que lhe concebe. Por exemplo, um colegiado formado majoritariamente por cientistas da computação que aprova a criação de um curso de SI em seu departamento tem maior probabilidade de produzir um currículo ancorado em aspectos técnicos; uma faculdade de Administração, terá maior probabilidade de produzir um currículo ancorado em aspectos procedimentais. As três possibilidades, associadas aos três aspectos atribuindo-lhe maior peso, são válidas e consistentes com a essência de SI. Propomos analisar a história e currículo de diversos cursos de SI espalhados pelo Brasil para ver como os mesmos podem diferenciar substancialmente, oficialmente aprovados como SI.

Por outro lado, pesquisas atravessando Ética e SI ainda são escassas [Bock et al. 2021]. No cenário brasileiro, a principal conferência dedicada ao pensar SI, o Simpósio Brasileiro de Sistemas de Informação (SBSI), reforça esta escassez e abandono do tema, sem trabalhos dedicados primariamente ao tópico e com uma deficiência alarmante de aspectos éticos entremeados, de forma não-primária (como ocorrência de Comitê de Ética) [Carvalho et al. 2021c].

6.4.1. Plágio

Plágio é a “ação ou efeito de plagiar, de expor ou de mostrar uma obra intelectual de outra pessoa como se fosse de sua própria autoria”, ou ainda, “apresentação que alguém faz de algo, como se fosse de própria autoria, quando na verdade foi criado ou pertence a outrem; cópia, imitação: plágio de um livro, trecho de música, de uma pintura etc”³⁸. O plágio é uma das mais importantes questões éticas no meio acadêmico, e portanto, em pesquisas em Sistemas de Informação.

Quando falamos em plágio na área acadêmica, normalmente nos referimos ao fato de alguém (professor, aluno, pesquisador) apresentar como de sua autoria, um texto, gráfico, imagem ou qualquer outro tipo de produção intelectual de outra pessoa, seja em todo ou em parte, sem o devido crédito. No caso do plágio acadêmico, existem diversos tipos. O plágio direto ocorre na forma da cópia literal do texto original, sem referência ao autor e sem indicar que é uma citação. Já o plágio indireto é uma reprodução, com as próprias palavras, das ideias de um texto original sem a devida indicação da fonte. Um caso bastante discutido também é o chamado auto-plágio, que ocorre quando um autor reapresenta, como se fosse original, um trabalho ou parte significativa de um trabalho, de

³⁸<https://www.dicio.com.br/plagio/>. Disponível em: 01/09/2021

sua própria autoria. Apesar de mais difícil de ser percebido, além desses tipos, também pode ser considerado plágio o uso de uma ideia com elementos muito parecidos com o original, ou seja, plágio de um conceito.

Além do dilema ético envolvido, plágio é um crime previsto em leis. Por exemplo, a violação dos direitos autorais prevista no artigo 184 do Código Penal ³⁹. O sujeito que comete plágio também pode ter que se retratar em público; e, por exemplo, no caso de um aluno, o ato pode levar à reprovação ou desligamento da instituição, e no caso de pesquisadores, inclusive professores, pode levar à demissão.

A Cartilha Sobre Plágio Acadêmico da UFF, disponível online ⁴⁰, apresenta detalhes sobre esse problema ético de forma clara e concisa, e dá a dica fundamental de como evitá-lo: “[...] basta escrever com suas próprias palavras de modo a explicar todas as citações, apresentar as fontes no próprio texto, e, se necessário, incluir as citações diretas (texto literal do autor utilizado) à medida que o trabalho vai sendo desenvolvido”. Desta forma, ao citar corretamente, você não vai cometer plágio. É importante também ressaltar que existem normas, tais como, NBR 10520:2002 e NBR 6023:2002, da Associação Brasileira de Normas Técnicas (ABNT) ⁴¹, que indicam as formas corretas que realizar as citações.

Plágio não está restrito a texto, figuras ou tabelas. Há também o plágio de código de software [Đurić e Gasevic 2013], tanto em ambiente acadêmico formal quanto científico, i.e., trabalhos acadêmicos ou artigos científicos. Por exemplo, um aluno desenvolve um software para uma disciplina e um ano depois cede este software para um colega utilizar o mesmo, idêntico, o que configura plágio; cientificamente, apropriar-se de código de software de outrem sem a devida citação e referência e publicar comunicação científica a partir do uso do mesmo, reivindicando autoria original para si. Bases de dados também são passíveis de plágio [Nunes e Pinheiro-Machado 2017], sem a devida citação e referência de uma base específica por culminar em conflito ou disputa judicial. Não há um costume de registrar código de software ou bases de dados para formalização de autoria, principalmente no meio da Computação. Podemos conjecturar que isto se dê porque na formação em Computação este tópico é ignorado, posto de lado ou considerado irrelevante de obrigatoriedade curricular; e que não haja uma cultura consolidada sobre estes elementos, principalmente no meio acadêmico.

Com a facilidade propiciada pela Computação, dados ou informações explicitamente estruturados estão amplamente disponíveis, basta um “Ctrl + c, Ctrl + c” (copiar e colar) para apropriar-se do texto de outros autores, ou do próprio. Ao mencionarmos a complexidade e heterogeneidade da Ética, ressaltamos que nem sempre o plágio é “individual” e que ocorre em uma cadeia de responsabilizações. Por exemplo, um orientador acadêmico que permite que um orientado siga para uma banca de defesa de monografia sem debate ou preocupação relacionada à plágio; uma conferência que permite trabalhos apresentando plágio explícito e não promove auto-reflexão sobre o fenômeno; um professor que descobre que um grupo de estudantes comete plágio entre si, e ignora ou não interfere educacionalmente.

³⁹http://www.planalto.gov.br/ccivil_03/leis/2003/110.695.htm. Disponível em: 01/09/2021

⁴⁰<http://www.noticias.uff.br/arquivos/cartilha-sobre-plagio-academico.pdf>. Disponível em: 01/09/2021

⁴¹<http://www.abnt.org.br>. Disponível em: 01/09/2021

6.4.2. Ciência Aberta e Transparência

Seguiremos as obras de [Pontika et al. 2015] e [Elliott 2020] nestes tópicos. Apesar da predominância de ideias positivas sobre estes tópicos, a deliberação ética deve ultrapassar uma agenda enviesada. Por exemplo, Transparência, assim como Ética, é um dos termos associados a valores clichês que todos os indivíduos ou organizações dizem apoiar ou defender simbolicamente, só que concretamente pouquíssimos apoiam ou investem ativamente, limitados ao campo discursivo de moralismo imaginário. Habilitar e gerenciar transparência culmina em consequências positivas e, contrariando o senso comum, negativas. Obviamente, Transparência é um dos requisitos básicos para uma sociedade democrática, só que ela não vem sem custos, para implantação ou consequenciais. Como indicam [Bannister e Connolly 2011], iniciativas de transparência resultam em aumento do escrutínio crítico em relação às atividades ou dados, permissibilidade para auditoria enviesada ou inescrupulosa, e abertura de dados que podem servir para agendas políticas escusas.

Um exemplo baseado na vivência profissional do primeiro autor. Em determinado momento participei de uma comissão dedicada à transparência na Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e em certa reunião foi exposto um caso associado com um dado que obrigatoriamente deve ser aberto, a agenda do Reitor. Antes, a agenda do Reitor era integralmente transparente no devido sítio eletrônico. Em um período de perturbação política manifestantes rastream a agenda do Reitor, identificando hora e local, e lançaram ovos no carro oficial do mesmo. Nos ausentamos da análise ética deste episódio específico, pois foge do domínio de SI primariamente. Todavia, os dados transparentes foram reavaliados, e a agenda do Reitor deixou de expor hora e local, indicando apenas o compromisso e respectiva parte do dia. Este é um exemplo relativamente inofensivo e ilustra como o mito da “transparência absoluta de tudo a todo momento” nem sempre é positivo.

A Figura 6.3 expõe uma taxonomia de Ciência Aberta de forma panorâmica. Observando-a podemos notar uma pluralidade de dados e informações associados com o conceito, agenda e agência da Ciência Aberta. Dados Abertos são essenciais ao pensar SI e facilmente associáveis com a prática de pesquisa, assim como habilitados pela autonomia de poder e capacidade do pesquisador. Por exemplo, enquanto o pesquisador nem sempre tem a capacidade financeira de publicar uma comunicação científica como acesso aberto, dependendo do preço, ele pode ter sua base de dados aberta online referenciando a mesma no próprio trabalho, por si.

Na Figura 6.4, complementada pela Tabela 6.4.2, é exposta uma taxonomia para transparência científica. **Propósito** responde **por que** a transparência está sendo habilitada; **Audiência** responde **quem** receberá os dados e informações transparentes; **Conteúdo** apresenta **o que** será transparente. Apresentando **Como**, temos **Prazo** apresentando as diversas opções temporais envolvidas no processo científico; **Ator** são as partes possibilitadas a comunicar o conteúdo científico; **Mecanismo** identifica e esclarece informações que podem assumir uma ampla variedade de formas; **Local** são os diferentes ambientes quais a comunicação científica pode ser transmitida e compartilhada. Por fim, adaptamos uma das dimensões de [Elliott 2020], e o que era “perigo” representamos como **Risco**. Este último trata das configurações ou consequências das iniciativas de transparên-

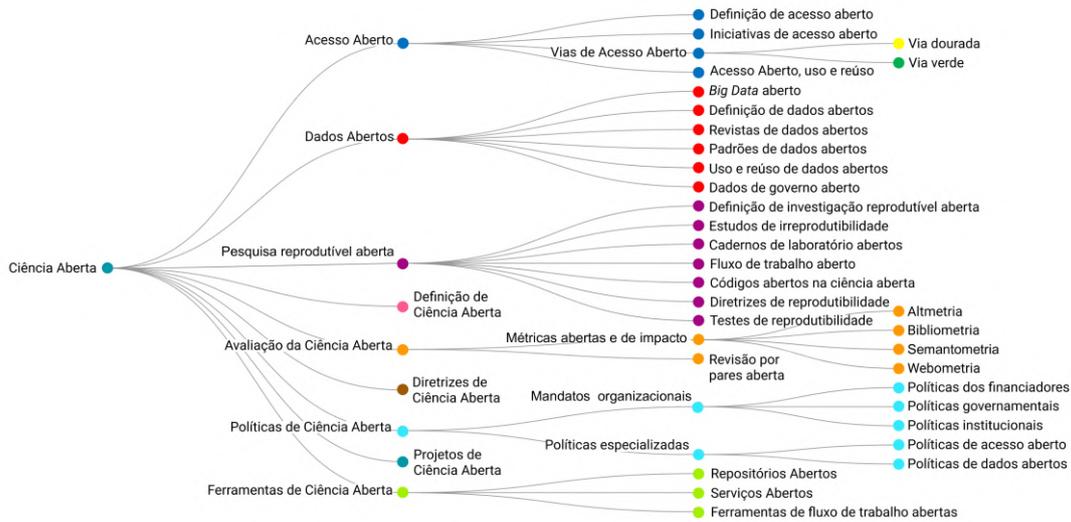


Figure 6.3. Taxonomia da Ciência Aberta [adaptado de [Pontika et al. 2015]]

cia que informam ou implementam as demais. Adaptamos para risco pois toda pesquisa científica apresenta riscos e perigo é (i) um termo “inflamado”, nem sempre um perigo é interpretado como tal e inconscientemente a ideia de risco é inerente; (ii) presume má fé, rotular uma comunicação científica como “perigosa” estreita a abertura de ponderação para intensidade, i.e., uma pesquisa pode ser mais ou menos arriscada, enquanto para inaceitável ser mais ou menos “perigosa”.



Figure 6.4. Representação de múltiplas dimensões nas quais as iniciativas de transparência podem variar, organizadas de acordo com quatro perguntas [adaptado de [Elliott 2020]]

Cada uma destas dimensões tem especificações extensas, detalhadas em [Elliott 2020]. Cada uma das dimensões e seus sub-itens podem ter desdobramentos próprios, como Audiência. Para comunicar-se com cada audiência específica há especificidades próprias, e uma forma de estabelecer este contato é utilizando Linguagens Cidadãs ou Claras, como para comunicar processos ao público-geral [Carvalho et al. 2020] ou problemáticas na comunicação gráfica de dados sobre COVID-19 [Oliveira et al. 2021].

Entretanto, a ideia de Ciência Aberta e Transparência deixa de ser trivial quando

passa da ideia à práxis. A maioria dos autores, incluídos [Pontika et al. 2015] e [Elliott 2020], indicam definições, bases, conceitos e ideias sobre estes temas, entretanto a análise materialista é vazia ou quase nula. Isto é, o que deve ser feito; quais atores devem ser manobrados; quem tem a perder e a ganhar; os fenômenos negativos são restritos ou especiais à Ciência?; o ônus do problema é majoritariamente dos pesquisadores? Isto é, eles são relapsos, preguiçosos, apresentam interesses obscuros, dentre outros; o “fazer Ciência” mudou durante o Século XX ou se mantém ancorado na práxis do Século XVIII? E, principalmente, debater a moralidade acerca desta abertura de informações científicas.

É intuitivamente simples pleitear e defender a argumentação em favor da Ciência Aberta ou da Transparência Científica, principalmente quando a linha argumentativa envolve “progresso”, “democracia” ou termos socialmente polêmicos e complexos de se contra-argumentar sem recair na falácia do Apelo ao Absurdo (*Reduction ad absurdum*), i.e., a Ciência “é Aberta” ou “é Fechada”. Não há um meio termo ou o ônus da abertura científica recai majoritariamente sobre os pesquisadores individualizados, ou grupos de pesquisa/laboratórios isolados.

Vamos citar apenas um exemplo, o preço para se publicar uma pesquisa em uma comunicação científica considerada de alta qualidade no Brasil. No Brasil os pesquisadores são pontuados e categorizados, de quantificação à qualificação, através do sistema de avaliação QUALIS ⁴². O capital moral e social do(s) pesquisador(es) é relacionado à sua produção e publicação de pesquisas nas melhores posições do QUALIS. O custo para publicar nos repositórios melhores qualificados pelo QUALIS variam entre US\$2000 e US\$3000 ⁴³. Considerada a situação atual da cotação do dólar, flutuando em aproximadamente R\$5,00, uma publicação pode custar entre R\$10000 e R\$15000, **em média**. Um pesquisador arca com a maior parte das despesas, investimentos ou gastos, da sua produção científica, e além deste valor isolado há o gasto com viagem, estadia, inscrição em conferência, alimentação, transporte, isso considerando apenas o escopo acadêmico-científico. Excluindo conferências? Sem problemas, temos os custos da edição, publicação, publicidade e divulgação científica de seu livro ou outra obra qualquer.

E considerando que as conferências e periódicos de melhor colocação no QUALIS são internacionais. E que os preços são em dólar. E que o dólar está, atualmente, cotado na flutuação próxima de R\$5,00. Isso sem considerar custos secundários da prática acadêmica, que são ainda mais opacos ao sistema, como “jantar da conferência”, transportes com certa qualidade para movimentação com equipamentos caros, colaboração financeira entre colegas, contratação de revisores para ajustes linguísticos, e reservas para quaisquer riscos. Transparência Científica é igualmente uma ameaça de tiro no pé e objeto de manobra para disputa ideológica econômico-política, por exemplo, por que pagar por “jantar da conferência”? Se é funcionário público, que use transporte público? Ao invés de se hospedar em hotel, que tal se hospedar em um *hostel*? Ao invés de viajar de avião, por que não suportar mais de 24 horas de viagem intermunicipal de ônibus?

Em suma, é óbvio que quando o debate recai em uma simplificação minimalista, será simples apontar “certo” e “errado” e similares. Entretanto podemos contra-

⁴²<https://pt.wikipedia.org/wiki/Qualis>

⁴³<https://www.openaccess.cam.ac.uk/publishing-open-access/how-much-do-publishers-charge-open-access>. Disponível em: 01/09/2021

Table 6.1. Oito dimensões da transparência científica [adaptado de [Elliott 2020]]

Propósito	Audiência
Facilitar a reanálise de resultados	Cientistas envolvidos na pesquisa
Melhorar a replicabilidade	Outros cientistas
Promoção da inovação	Outros acadêmicos
Manter a responsabilidade dos especialistas	Desenvolvedores de políticas
Facilitar a interação crítica	Políticos e entidades políticas
Promover a formulação de políticas de alta qualidade	Jornalistas
Capacitar o público a tomar decisões de acordo com seus valores	Grupos de partes interessadas específicas
Promover confiabilidade	Público geral

Conteúdo	Prazo
Dados, métodos, código, materiais	Antes do início da pesquisa
Interpretações dos dados, métodos e código para não especialistas	Ao longo do processo de pesquisa
Julgamentos de valor de vários tipos	Imediatamente após os dados serem coletados
Valores ou fatores que influenciam os julgamentos	Após a publicação
Relatórios de deliberações subjacentes	Durante ou após revisões ou análises subsequentes da pesquisa
Implicações de julgamentos de valor	

Ator	Mecanismo
Cientistas que realizaram a pesquisa	Discussões entre cientistas (oral ou escrita)
Outros cientistas	Colaborações interdisciplinares
Acadêmicos trabalhando em outras áreas	Colaborações com membros da comunidade
Jornalistas	Órgãos consultivos do governo e outras iniciativas
Sociedades científicas	Processos contraditórios
Agências governamentais	
Organizações não governamentais e organizações da sociedade civil	

Local	Risco
Comunicação por cientistas (oral ou escrita, incluindo mídia social)	Desperdiçar recursos escassos
Registros e repositórios	Desacelerando a ciência
Jornalismo científico	Prejudicar empresas
Relatórios de agências governamentais	Violar privacidade
Relatórios de organizações não governamentais ou grupos comunitários	Gerando ceticismo impróprio
	Criar uma falsa sensação de confiança
	Causando confusão
	Facilitar esforços para assediar ou enganar

argumentar tanto deontológica quanto teleologicamente, nenhum destes dois tópicos aqui presentes são tão simples quanto parecem.

6.4.3. Consentimento, Assentimento e Comitê de Ética

De modo a resguardar a integridade dos participantes de pesquisas científicas, a Comissão Nacional de Ética em Pesquisa (CONEP), vinculada ao Ministério da Saúde, determina várias Normas que devem ser seguidas pelas pesquisas que envolvem seres humanos. A CONEP é responsável por elaborar e atualizar as diretrizes e normas para a proteção dos sujeitos da pesquisa; avaliar e acompanhar os protocolos de pesquisa em áreas temáticas especiais e coordenar a rede de Comitês de Ética em Pesquisa das instituições [Cosac 2017]. Cada instituição pode ter um ou mais Comitês de Ética em Pesquisa (CEP), que é um colegiado interdisciplinar e independente, que delibera sobre a aprovação de projetos de pesquisa e tem como meta defender os interesses dos participantes de pesquisas, considerando sua integridade e dignidade.

Os Comitês de Ética em Pesquisa têm a sua origem justificada pelas atrocidades

cometidas nas guerras do século passado. Logo, a princípio, as normas e diretrizes eram focadas em pesquisas da área biomédica. O Conselho Nacional de Saúde (CNS) lançou uma primeira Resolução em 1988. Ao longo desses primeiros anos o processo de submissão e avaliação de projetos foram se aprimorando e surgiram necessidades não supridas por essa resolução. Então, em 1996 foi lançada uma nova resolução denominada Resolução CNS no. 196/1996, que propunha como diferencial em relação à anterior quatro referenciais básicos da bioética: autonomia, não maleficência, beneficência e justiça, assegurando direitos e deveres dos pesquisadores e dos participantes das pesquisas, considerando fortemente aspectos ligados a pesquisas da área biomédica. E em 2012, essa resolução foi aperfeiçoada, tendo sido lançada a Resolução CNS no. 466/2012 [Brasil 2012].

Mas, com o passar do tempo e com a difusão de pesquisas em muitas áreas humanas e sociais, a atuação dos CEPs foram se ampliando e grupos dessas áreas iniciaram uma grande discussão nacional, que culminou na proposta de uma nova norma, que abarca pesquisas em áreas além da biomédica, a Resolução CNS no. 510/2016 [Brasil 2016b]. A Resolução 510/2016 tem como característica diferencial em relação à Resolução 466/2012 a não intervenção direta no corpo humano. Outro aspecto a destacar é a possibilidade de obter o consentimento para participação nas pesquisas de forma oral, ou por imagem.

E as pesquisas da área de SI, em que resolução se encaixam? A princípio pesquisas de testes de software, avaliações de usabilidade e da qualidade de interface, e testes de produtos de software para contextos educacionais se conectam mais fortemente com a Resolução 510/2016. Entretanto, se a pesquisa tiver participação de menores de 18 anos ou de pessoas com deficiência cognitiva, alguns aspectos específicos devem ser considerados. Neste caso, será necessária a autorização de um responsável.

Em março de 2018 foi publicada a Resolução CNS no. 580/2018 que trata de pesquisas que usam o Sistema Único de Saúde (SUS) [Brasil 2018b]. Em 2021, motivado pela pandemia de COVID-19, a CONEP publicou um Ofício ⁴⁴ para orientar pesquisadores e CEPs em relação a procedimentos que envolvam o contato com participantes e/ou coleta de dados em ambiente virtual.

A submissão de projetos ao CEP requer o preparo de vários documentos, para que os riscos éticos possam ser avaliados. A submissão é realizada na Plataforma Brasil ⁴⁵ e vários documentos devem ser preparados considerando as resoluções vigentes. Para a realização do cadastro na plataforma é preciso ter a cópia de documento de identificação escaneado e o link do currículo LATTES.

O projeto deve ter informações que permitam avaliar possíveis riscos aos participantes e documentos anexos, que contemplam vários aspectos das resoluções associadas. De modo a ajudar na submissão do projeto na plataforma, alguns aspectos devem ser descritos no projeto: Resumo e Desenho da Pesquisa; Contextualização e objetivos; Definição da população e como será contatada; Método de coleta de dados; Riscos e benefícios; Orçamento; Instrumento de coleta de dados e Cronograma. Ressaltamos que não

⁴⁴http://conselho.saude.gov.br/images/Oficio_Circular_2_24fev2021.pdf. Disponível em: 01/09/2021

⁴⁵<https://plataformabrasil.saude.gov.br/login.jsf>. Disponível em: 01/09/2021

são avaliados projetos já iniciados, ou realizados.

De acordo com a Resolução 466/2012 [Brasil 2012], o assentimento livre e esclarecido é voltado para o registro da anuência do participante da pesquisa, sendo criança, adolescente ou legalmente incapaz, que devem ser esclarecidos sobre a natureza da pesquisa, seus objetivos, métodos, benefícios previstos, potenciais riscos, considerando seu nível de compreensão e idade. Esta anuência é dada em um documento separado do Termo de Consentimento Livre e Esclarecido (TCLE) dos responsáveis. Logo, os pais/responsáveis assinam o TCLE, consentindo a participação de menores de idade ou pessoas com deficiência.

Os documentos a serem anexados na submissão são: TCLE (se for consulta a banco de dados é possível anexar um documento pedindo a dispensa); Termo de Consentimento do responsável (quando o participante for menor de idade); Termo de Assentimento Livre e Esclarecido (TALE) (para menor de 12-17 anos). Para menores de 12 anos, somente o responsável precisa assinar, mas não há impedimento das crianças também assinarem um documento de assentimento. Termo de Anuência da Instituição onde a pesquisa será realizada (TAI) com assinatura do responsável pelo setor onde a pesquisa se desenvolverá; Folha de rosto, devidamente preenchida e assinada pela direção, ou coordenação do local de vinculação do pesquisador principal. Em geral, para evitar conflitos de interesse: se o coordenador ou o diretor forem da equipe do projeto, eles não podem assinar a Folha de Rosto. Da mesma forma, o Termo de Anuência Institucional não pode ser assinado por pesquisador membro da equipe do projeto.

O TCLE deve ser escrito em forma de convite, com linguagem acessível ao seu público-alvo. Deve conter o objetivo; como será realizada a pesquisa; a possibilidade de desistência a qualquer momento, sem que cause prejuízo para o participante; os riscos e os benefícios; a não-remuneração; garantia do sigilo dos resultados; como serão publicados os resultados; a garantia do sigilo; o contato dos pesquisadores, o contato do CEP e os espaços para as assinaturas. Toda a pesquisa oferece riscos mínimos e isso deve estar explícito no projeto e no TCLE. Neste caso, pode ser um constrangimento, ou uma recuperação de memórias desagradáveis, manchas no braço por coleta de sangue, dentre outros. O TCLE deve conter claramente os procedimentos para minimizar os riscos. As filmagens devem ser explicitamente citadas no TCLE. E se forem realizadas para coleta de dados e transcrição, o texto explicativo deve estar no corpo do TCLE. Filmagens para exibição pública devem constar em parágrafo separado. Em geral, as instituições têm modelos da documentação em seus sites.

6.5. Sobre os extremos e seus valores

Do parcial ao imparcial, do neutro ao tendencioso, do enviesado ao sem viés, do isento ao comprometido, do ruidoso ao sem ruído, daquele com valor ao sem valor, do autônomo ao dependente... e as muitas mais variações deste binarismo irreal. Como autor central deste debate, recorreremos à [Simon 2019] para dialogar com as Ciências do Artificial. Tecnologias alteram a realidade e a nossa realidade altera as tecnologias. Nós fazemos parte da realidade, tecnologias nos alteram e nós alteramos tecnologias. Todos somos dotados de valores morais, nossos valores morais alteram a realidade e a realidade altera nossos valores morais. Isto é, nossos valores morais alteram tecnologias e tecnologias

alteram nossos valores morais. Esta relação dialógica ou dialética persiste há milênios, desde a ideação de tecnologias imateriais, como artefatos de comunicação.

O Painel Intergovernamental sobre Mudanças Climáticas (IPCC – *Intergovernmental Panel on Climate Change*) publicou um relatório em agosto de 2021 apontando que a humanidade está sendo conduzida à extinção através das mudanças climáticas, caso haja omissão de ações práticas para significativamente mitigá-la ou controlá-la. A responsabilidade principal é dos seres humanos ⁴⁶, enquanto diversos efeitos são inevitáveis e irreversíveis ⁴⁷. Dados e gráficos apontam que o arrefecimento deste fenômeno piora exponencialmente a partir de meados do Século XIX, coincidindo com a Revolução Industrial e proliferação tecnológica de produções massivas. É inegável de que os avanços tecnológicos alteraram substancialmente a vida dos seres humanos, positivamente em diversos aspectos, entretanto em outros aspectos estamos sendo conduzidos ao nosso próprio holocausto. O Brasil, por exemplo, parece estar indo na contramão dos esforços para melhoria climática ⁴⁸.

Podemos falar de plágio, normas jurídicas ou técnicas, comitês de éticas, ciência aberta e ainda outros muitos pontos associados à SI, entretanto, precisamos refletir na nossa responsabilidade para com a sustentabilidade da nossa região, do nosso país e do nosso planeta. Seria hipocrisia da nossa parte ignorar o macro-estrutural, porque foi justamente esta negligência macro-estrutural que contribuiu para uma pandemia de COVID-19.

Em um exemplo que une Ciência Aberta, divulgação científica, para instrução climática, o episódio “Poeira” (E03) ⁴⁹, da série “A Era dos Dados” (T01), expõe a relação entre a poeira que sai do deserto do Saara e chega à Amazônia, 27 bilhões de toneladas por ano. A poeira que sai do Chade, na África, tem função nutritiva para a vegetação na Bacia Amazônica. Milhares de toneladas de Fósforo são carregadas pelos ventos e ajudam no balanceamento ecossistêmico da floresta, a poeira traz elementos que reabastecem o que a chuva leva. Entretanto, isto não é tecnologia, não “é ético” e nem passível de moralidade, os ventos não escolhem interagir positivamente com a Bacia Amazônica, o deserto não escolheu ceder sua poeira e a Bacia Amazônica não demandou por estes recursos diretamente de uma região africana separada por um imenso oceano. Por outro lado, o desmatamento humano desenfreado está conduzindo a Floresta Amazônica ao ponto de “não retorno” ⁵⁰, nem a floresta, nem os ventos e nem o deserto africano tem participação moral ou ética neste cenário, apenas alguns seres humanos inescrupulosos e enviesados em favor desta prática. Desmatar a Amazônia é uma escolha ética, e apesar dos pareceres alarmantes seguidos, os atores que o fazem, e reforçam esta prática, seja simbólica ou concretamente, escolheram manter este curso de ação.

Você não estaria lendo este texto se não fosse pela Computação; cientistas não

⁴⁶<https://www.theguardian.com/environment/2021/aug/09/climate-crisis-unequivocally-caused-by-human-activities-says-ipcc-report>. Disponível em: 09/01/2021

⁴⁷<https://www.theguardian.com/science/2021/aug/09/humans-have-caused-unprecedented-and-irreversible-change-to-climate-scientists-warn>. Disponível em: 09/01/2021

⁴⁸<https://www1.folha.uol.com.br/colunas/helio-mattar/2021/08/relatorio-mostra-que-o-brasil-caminha-na-contramao-do-mundo.shtml>. Disponível em: 01/09/2021

⁴⁹https://www.imdb.com/title/tt12816820/?ref_=ttep_ep3. Disponível em: 01/09/2021

⁵⁰<https://www.bbc.com/portuguese/brasil-5046861>. Disponível em 01/09/2021

teriam estudado o fluxo da poeira de forma precisa se não fosse pela Computação; a divulgação científica agenciada pela *Netflix* não seria possível sem a Computação; a análise de que estamos chegando ao ponto de “não retorno” florestal não seria tão acurado sem a Computação; as análises científica do IPCC não seriam possíveis sem a Computação. Por outro lado, toda a aceleração da destruição amazônica, escolha de localizações estratégicas para extração e comunicação das redes de desmatamento também não seria possível sem a Computação. Todos estes elementos são sistemas de informação.

Esclarecemos que esta instância serve como ilustração apenas e nossa intenção se afasta de uma agenda de urgência para que toda comunidade de Computação foque nas mudanças climáticas. O objetivo é mostrar do local ao macro-estrutural, do natural ao artificial, da complexidade e heterogeneidade dos valores e, por fim, da conscientização do poder prático que a Computação tem, seja no pensar ou no fazer. E quem determina a maioria das nossas próprias diretrizes éticas somos nós mesmos, através das tecnologias que nos cercam, e dos valores morais que são influenciados por elas e as influenciam.

Interagimos com a realidade baseados nas limitações das nossas subjetividades, nossas vivências, experiências, ideologias, identidades, limites cognitivos, conhecimentos, dentre outros. Assim como a realidade incide sobre nossas subjetividades. Simultaneamente, conforme aumentamos o grau de complexidade analítica da realidade, se torna menos categórica a valoração moral, ao ponto que deixamos de analisar “o que?” e partimos outros questionamentos como “para quem?”, “como?”, “quando?”, “por que?”. Por exemplo, [Bender et al. 2021] levanta o questionamento do impacto sustentável que IAs colossais com milhões de linhas de código, como BERT, causam e as possíveis consequências negativas; como sustentabilidade natural e social. Uma das autoras de [Bender et al. 2021], Timnit Gebru, foi abruptamente desligada da empresa Google e uma explicação possível trata do seu envolvimento neste artigo crítico “atacando” IAs colossais, categoria na qual BERT se enquadra ⁵¹. BERT faz parte do motor das pesquisas do Google, e foi parte determinante do lucro de US\$26 bilhões de dólares da empresa no terceiro quartil do ano. Isto é, os benefícios de IAs colossais suplantam seus possíveis reveses? Ou as mesmas são apenas ferramentas de lucratividades desenfreada? Com o cenário de mudanças climáticas atual, mesmo assim vale a pena? Por exemplo, qual impacto que BERT tem nas mudanças climáticas?

Cabe a nós, enquanto indivíduos ou coletividade organizada, reconhecer que a imparcialidade é um mito e que, apesar de almejarmos buscá-la, a parcialidade faz parte do ser humano, assim como das tecnologias e artefatos que desenvolvemos. Concretamente, inexistente algo totalmente bom ou mau, certo ou errado, justo ou injusto, e cabe a nós calibrar as análises a partir de ponderação ética consciente e racional. A totalidade cobre elementos diversos externos ao objeto em si, como público-alvo e possíveis consequências. Precisamos pensar nas responsabilidades e valores das nossas ações, seja do pensar ou do fazer [Lacey 2019] e das tecnologias e artefatos que idealizamos explicitamente, projetamos e implantamos [Shilton 2018].

⁵¹ <https://www.theverge.com/2020/12/5/22155985/paper-timnit-gebru-fired-google-large-language-models-search-ai>. Disponível em: 01/09/2021

6.5.1. A história que se repete...

[Silva 2020] categoriza o infame caso onde “Google Photos marca fotos de jovens negros com a tag ‘Gorila’”, como uma microagressão de negação de cidadania e microinsulto. Este caso ocorreu em 2015. Em Setembro de 2021, o algoritmo do *Facebook* novamente associa pessoas pretas à “primatas”⁵², como ilustrado na Figura 6.5, o vídeo expõe uma pessoa preta e a plataforma questiona “Continuar vendo vídeos sobre Primatas?”.



Figure 6.5. Caso explícito de racismo na plataforma *Facebook*

Rapidamente um porta-voz da empresa surgiu para pedir desculpas, e reiterar os esforços da organização para “prevenir que casos assim ocorram novamente” e que este é um “erro inaceitável”. Darci Groves, uma ex-funcionária que foi gestora de conteúdo na empresa, “sugeriu que lidar com problemas de raça não é uma prioridade para os líderes [na empresa Facebook]”⁵³.

Analisando reações na própria plataforma *Facebook*, comentários são diversos, muitos com um posicionamento “defensivo” favorável à empresa ou à IA. Por exemplo, que nós todos realmente somos primatas; que a IA “aprendeu” com as interações de dados entre os usuários; que a IA é “lógica” e imparcial; que racistas são os usuários, não a IA; e acusações aleatórias entre usuários que não sabem o que é ou como funciona uma IA. [Benjamin 2019] apresenta um viés crítico sobre este tema, indo contra esta linha argumentativa de que todo o agenciamento do fenômeno racista tem o seu ônus na ação dos usuários, i.e., ele é um fato social em conformidade; e que o controlador intermediário é apenas um “facilitador” sem responsabilidade efetiva. Incluindo respaldos em sustentabilidade, como sustentabilidade social neste caso.

⁵²<https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html>. Disponível em 01/09/2021

⁵³<https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html>. Disponível em: 03/09/2021

Um dilema ético que necessita de superação nas RSO trata da responsabilização entre usuários, incluídos coletivos e organizações; a plataforma, o sistema em si; e as partes responsáveis, gestoras do sistema.

Após um dos casos mais escandalosos e imorais relacionados com desinformação e SI, a história se repete mais uma vez com uma nova roupagem, mas na mesma plataforma, com o mesmo corpo de líderes e fenômenos significativamente similares. O escândalo Cambridge Analytica, relacionado diretamente às eleições presidenciais nos EUA de 2016, foi um marco exemplar para Ética Computacional [Spinello 2020]. Novamente, agora nas eleições presidenciais de 2020, desinformações inundaram o *Facebook* e receberam seis vezes mais cliques se comparadas com informações factuais⁵⁴, e mesmo que este fenômeno apresenta maior intensidade quando relacionado a usuários categorizados como de Direta, ocorre também com aqueles de Esquerda. Em resposta, a empresa Facebook respondeu que este quantitativo reflete apenas cliques, não quantas pessoas realmente viram o conteúdo. A quantidade de pessoas que, efetivamente, veem o conteúdo não é transparente pela empresa.

Baseando-se em brechas interpretativas de seus próprios termos de uso, Facebook impede pesquisadores de realizar pesquisas com maior detalhamento e profundidade sobre os fenômenos de desinformação em sua plataforma⁵⁵, bloqueia contas, cerceia acesso a dados, e apela a brechas contratuais. Ao serem impedidos de pesquisar, os pesquisadores alegam que a empresa busca formas de pará-los de analisar os fenômenos recorrentes em sua principal plataforma.

Concluindo, casos persistem e fenômenos se repetem. Pouco é efetivamente feito. Reforçamos que o pensar e o fazer devem seguir de mãos dadas, e como percebemos nestes breves exemplos, quem está fazendo está ignorando o que está sendo pensado, e quem está pensando está ingenuamente acreditando que o fazer irá simplesmente ou prontamente se adequar.

6.6. Conclusão?

Nesta obra cobrimos diversos pontos atravessados entre Ética, Computação, SI e Pesquisa. Todavia, eles foram um pingô no oceano de conteúdo que este tema oferece, e que abre lacunas, espaços, vias e um universo de possibilidades para análises profundas, ricas e que necessitam de qualidade ímpar. E esta qualidade não se esgota na Computação, porque a Computação é uma extensão da prática humana em um nível de refinamento e complexidade pouco intuitivo, e até um pouco perverso. Qual o nosso problema neste ponto? Construir um *corpus* de conhecimento para reflexão ética que habilite uma conscientização ética panorâmica, holística e *epistemológica* [Escobar 2018] nos contextos, meios e cenários brasileiros. Dito isto, este trabalho não se conclui, e se lhe fizemos refletir, agora a “batata-quente” está com você. Nenhuma obra da Ética se conclui, porque a própria ideia de “concluir” um chamado à Ética é, por si só, absurdo à própria Ética.

Sobre o fazer SI, analisamos o quão pervasivo o *WhatsApp* está se tornando na

⁵⁴<https://www.washingtonpost.com/technology/2021/09/03/facebook-misinformation-nyu-study/>. Disponível em: 03/09/2021

⁵⁵<https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html>. Disponível em: 01/09/2021

realidade brasileira dos negócios; criticamos o estado discursivo de que uma Ética para SI está pautada em organizações privadas ou empresas, colocando organizações públicas e organizações semi-formais em segundo plano (ou em plano nenhum, descartando-as); trouxemos alguns exemplos de Códigos de Ética ou de Conduta, mostrando que os acordos morais variam entre coletividades; debatemos sobre normas jurídicas e técnicas, e pela própria previsão de [Moor 1985] a tendência é que nestes próximos anos a formulação de leis relacionadas à Computação, ou seu uso e aplicação, cresça, e que nem sempre estas leis considerarão o estado da moralidade ou apresentarão deliberação ética racional coerente; com a Computação “invadindo” todos os possíveis espaços, diversas variáveis entram na equação da tomada de decisão ética ao se tratar dos valores e das consequências associados às soluções computacionais.

Trouxemos a importância do atravessamento entre Ética e pesquisa em SI sobre o pensar SI, acima disso, do respeito e consideração à interdisciplinaridade da área. Plágio não se resume à texto, imagens ou tabelas, se estende para códigos de software e bases de dados; a Ciência Aberta e Transparência Científica são “objetivos” para uma democratização e abertura de diálogo ampla tanto entre estudiosos, cientistas ou o público externo à Academia, todavia tem seus percalços, como preços e disputas econômico-políticas, que apesar de saudáveis em um ambiente democrático podem ser enviesadas ao absurdo; e apontamos a importância de se respeitar os participantes das pesquisas, considerando sua identidade, solicitando assentimento às crianças e consentimento aos adultos, e envolvendo comitês de ética nas pesquisas computacionais com seres humanos participantes, alcançando uma validação institucional formal de que a respectiva prática científica teve deliberação ética ampla para ocorrer.

Por fim, estendemos o debate éticos aos extremos, à iminência do holocausto da raça humana, da complexidade sistêmica dos ecossistemas naturais e tecnológicos, apontando a importância, relevância e poder que a Computação oferece a todos nós, a comunidade e sociedade que nos cercam. Enquanto interessados no pensar-fazer computacional, principalmente de seus usos e aplicações pela área de SI, precisamos refletir com consciência ética sobre o nosso papel no todo, tanto das ações como inações, i.e., escolher não agir ou agir contra e combater.

Recorrendo a [Voltaire 2012]: a parcialidade não é uma condição agradável, mas a imparcialidade ou neutralidade são absurdas ⁵⁶.

Referências

- [Anderson et al. 1993] Anderson, R. E., Johnson, D. G., Gotterbarn, D., e Perrolle, J. (1993). Using the new acm code of ethics in decision making. *Commun. ACM*, 36(2):98–107.
- [Baase e Henry 2017] Baase, S. e Henry, T. M. (2017). *Gift of Fire, A: Social, Legal, and Ethical Issues for Computing Technology*. Pearson, New York, NY, 5ª edição.
- [Bannister e Connolly 2011] Bannister, F. e Connolly, R. (2011). The trouble with transparency: A critical review of openness in e-government. *Policy & Internet*, 3(1):1–30.

⁵⁶*Le doute n'est pas une état bien agréable, mais l'assurance est un état ridicule.* A dúvida não é uma condição agradável, mas a certeza é absurda.

- [Barger 2008] Barger, R. (2008). *Computer Ethics: A Case-Based Approach*. Cambridge University Press, Cambridge, RU.
- [Bender et al. 2021] Bender, E. M., Gebru, T., McMillan-Major, A., e Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA. Association for Computing Machinery.
- [Benjamin 2019] Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, New York, NY.
- [Bioni 2019] Bioni, B. (2019). *Proteção de Dados Pessoais A função e os limites do consentimento*. GEN, Barueri, Brasil.
- [Bobbio 2019] Bobbio, N. (2019). *Teoria do Ordenamento Jurídico*. Edipro, 2ª edição.
- [Bock et al. 2021] Bock, A., España, S., Gulden, J., Jahn, K., Nweke, L. O., e Richter, A. (2021). The ethics of information systems: The present state of the discussion and avenues for future work. Number 51 in *ECIS 2021 Research-in-Progress Papers*.
- [Brasil 2012] Brasil (2012). RESOLUÇÃO Nº 466, DE 12 DE DEZEMBRO DE 2012. Disponível em 01/09/2021 em: <https://cutt.ly/mmsS8Eua>.
- [Brasil 2016a] Brasil (2016a). RESOLUÇÃO Nº 5, DE 16 DE NOVEMBRO DE 2016. Disponível em 01/09/2021 em: <http://portal.mec.gov.br/docman/novembro-2016-pdf/52101-rces005-16-pdf/file>.
- [Brasil 2016b] Brasil (2016b). RESOLUÇÃO Nº 510, DE 7 DE ABRIL DE 2016. Disponível em 01/09/2021 em: <http://bit.ly/2fmnKeD>.
- [Brasil 2018a] Brasil (2018a). LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. Disponível em 01/09/2021 em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm.
- [Brasil 2018b] Brasil (2018b). RESOLUÇÃO Nº 580, DE 22 DE MARÇO DE 2018. Disponível em 01/09/2021 em: <https://conselho.saude.gov.br/resolucoes/2018/Reso580.pdf>.
- [Bynum 2000] Bynum, T. W. (2000). The foundation of computer ethics. *SIGCAS Comput. Soc.*, 30(2):6–13.
- [Bynum 2018] Bynum, T. W. (2018). Computer and Information Ethics. In *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*. <https://plato.stanford.edu/archives/sum2018/entries/ethics-computer/>. Disponível em: 01/09/2021.
- [Carvalho et al. 2020] Carvalho, L. P., Cappelli, C., e Santoro, F. M. (2020). BPMN Pra GERAL, Business Process Models in a Citizen Language. In *XVI Brazilian Symposium on Information Systems*, SBSI'20, New York, NY, USA. Association for Computing Machinery.

- [Carvalho et al. 2021a] Carvalho, L. P., Oliveira, J., e Santoro, F. M. (2021a). A PRESENÇA DE CONTEÚDOS SOBRE ÉTICA COMPUTACIONAL NA LITERACIA EM COMPUTAÇÃO INSTITUCIONAL BRASILEIRA. Encontro Virtual ABCiber 2021. (no prelo).
- [Carvalho et al. 2021b] Carvalho, L. P., Oliveira, J., e Santoro, F. M. (2021b). COMPUTAÇÃO, LITERACIA E ÉTICA COMPUTACIONAL. UM ESTUDO EXPLORATÓRIO PELO CIBERESPAÇO BRASILEIRO. Encontro Virtual ABCiber 2021. (no prelo).
- [Carvalho et al. 2021c] Carvalho, L. P., Suzano, J. A., Oliveira, J., e Santoro, F. (2021c). Ethics: What is the research scenario in the brazilian symposium sbisi? In *Anais da VII Escola Regional de Sistemas de Informação do Rio de Janeiro*, pp. 24–31, Porto Alegre, RS, Brasil. SBC.
- [Cesário et al. 2020] Cesário, J., Flauzino, V., e Mejia, J. (2020). Metodologia científica: Principais tipos de pesquisas e suas características. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, 5(11):23–33.
- [Cosac 2017] Cosac, D. C. d. S. (2017). Autonomia, consentimento e vulnerabilidade do participante de pesquisa clínica. *Revista Bioética*, 25:19–29.
- [Darch 2004] Darch, C. (2004). Digital divide or unequal exchange? how the northern intellectual property rights regime threatens the south. *International Journal of Legal Information*, 32(2):488–501.
- [de Pádua 2019] de Pádua, E. M. M. (2019). *Metodologia da pesquisa: abordagem teórico-prática*. Papirus Editora.
- [Deigh 2010] Deigh, J. (2010). *An Introduction to Ethics*. Cambridge University Press, Nova Iorque.
- [Dimoulis 2016] Dimoulis, D. (2016). *Manual de Introdução ao Estudo do Direito*. Revista dos Tribunais, 7ª edição.
- [Donia e Shaw 2021] Donia, J. e Shaw, J. (2021). Ethics and values in design: A structured review and theoretical critique. *Science and Engineering Ethics*, 27.
- [Elliott 2020] Elliott, K. C. (2020). A Taxonomy of Transparency in Science. *Canadian Journal of Philosophy*, page 1–14.
- [Escobar 2018] Escobar, A. (2018). *Designs for the Pluriverse Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press, Durham, USA.
- [Ferraz 2014] Ferraz, C. A. (2014). *Ética Elementos Básicos*. NEPFIL online, Pelotas, RS.
- [Fieser 2020] Fieser, J. (2020). Ethics. In *The Internet encyclopedia of philosophy*. <https://iep.utm.edu/ethics/>. Disponível em: 01/09/2021.

- [Garbe 1909] Garbe, W. (1909). Índios Botocudos : foto 04. 1 fotografia. 17 x 12 cm. Disponível em 01/09/2021 em: <https://brasilianafotografica.bn.gov.br/brasiliانا/handle/20.500.12156.1/570>.
- [Garon 2018] Garon, J. M. (2018). Reordering the chaos of the virtual arena: harmonizing law and framing collective bargaining for avatar actors and digital athletes. In Barfield, W. e Blitz, M. J., (eds), *Research Handbook on the Law of Virtual and Augmented Reality*, pp. 513—566. Edward Elgar Publishing, Cheltenham, Reino Unido.
- [Gibson 2014] Gibson, K. (2014). *An Introduction to Ethics*. PEARSON, Nova Jérsey.
- [Gyekye 2010] Gyekye, K. (2010). African Ethics. In *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*. <https://plato.stanford.edu/archives/fall2011/entries/african-ethics/>. Disponível em: 01/09/2021.
- [Johnson 2008] Johnson, D. (2008). *Computer Ethics*. Pearson, 4ª edição.
- [Kohun et al. 2012] Kohun, F., Rodi, A. F., e Delorenzo, G. (2012). Does information systems suffer an identity crisis? a case study of confusion and misinformation. *Issues in Information Systems*, 13(2):328–335.
- [Kon et al. 2020] Kon, F., Lago, N., Meirelles, P., e Denner, C. (2020). Direitos autorais, licenças e patentes. In Maciel, C. e Viterbo, J., (eds), *Computação e sociedade: a profissão - volume 1*, pp. 12–46. EdUFMT Digital, Mato Grosso, Brasil.
- [Lacey 2006] Lacey, H. (2006). Relações entre fato e valor. *Cadernos de Ciências Humanas - Especiaria*, 6(16):251–266.
- [Lacey 2019] Lacey, H. (2019). Adoção de medidas de precaução diante dos riscos no uso das inovações tecnocientíficas. *Estudos Avançados*, 33(95).
- [Laudon e Laudon 2020] Laudon, K. e Laudon, J. (2020). *Management Information Systems: Managing the Digital Firm*. Pearson, New York, NY, 16ª edição.
- [Marques 2016] Marques, I. C. (2016). História das ciências, estudos cts e os brasis. In *Abertura do Scientiarum Historia*.
- [Marques 2014] Marques, R. M. (2014). Contribuições À crítica da propriedade intelectual. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 7(2).
- [Mascaro 2013] Mascaro, A. (2013). *Estado e forma política*. Boitempo.
- [Masiero 2013] Masiero, P. C. (2013). *Ética em Computação*. EDUSP, São Paulo, SP.
- [Mingers e Walsham 2010] Mingers, J. e Walsham, G. (2010). Toward ethical information systems: The contribution of discourse ethics. *MISQ.*, 34(4):833–854.
- [Moor 1985] Moor, J. H. (1985). What is computer ethics?*. *Metaphilosophy*, 16(4):266–275.

- [Moriarty 2021] Moriarty, J. (2021). Business Ethics. In *The Stanford Encyclopedia of Philosophy (Summer 2021 Edition)*. <https://plato.stanford.edu/archives/sum2021/entries/ethics-business/>. Disponível em: 01/09/2021.
- [Nunes e Pinheiro-Machado 2017] Nunes, M. A. S. N. e Pinheiro-Machado, R. (2017). Propriedade intelectual e busca de informação tecnológica na área da computação. In Araújo, R. M. e Chueri, L. O. V., (eds), *Pesquisa e Inovação: Visões e Interseções*, pp. 67–92. PUBL!T, São Paulo, SP.
- [Oliveira et al. 2021] Oliveira, R., Cappelli, C., e Oliveira, J. (2021). An indicator of inefficient visualizations: The challenge of transparency during the covid-19 pandemic in brazil: An indicator of inefficient visualizations: The challenge of transparency during the covid-19 pandemic in brazil. In *XVII Brazilian Symposium on Information Systems, SBSI 2021*, New York, NY, USA. Association for Computing Machinery.
- [Padilha et al. 2005] Padilha, M. I. C., Ramos, F. R. S., Borenstein, M. S., e Martins, C. R. (2005). A responsabilidade do pesquisador ou sobre o que dizemos acerca da ética em pesquisa. 14(1):96–105.
- [Paradice et al. 2018] Paradice, D., Freeman, D., Hao, J., Lee, J., e (2018), D. H. (2018). A review of ethical issue considerations in the information systems research literature. *Foundations and Trends R in Information Systems: Vol. 2*, 2(2):117–236.
- [Pereira et al. 2015] Pereira, R., Baranauskas, M. C. C., e Liu, K. (2015). The value of values for hci: An informed discussion beyond philosophy. In *Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems, IHC '15*, New York, NY, USA. Association for Computing Machinery.
- [Perelman 2003] Perelman, M. (2003). *Steal This Idea: Intellectual Property and the Corporate Confiscation of Creativity*. Palgrave Macmillan.
- [Pontika et al. 2015] Pontika, N., Knoth, P., Cancellieri, M., e Pearce, S. (2015). Fostering open science to research using a taxonomy and an elearning portal. In *iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business*.
- [Ramalho 2013] Ramalho, R. A. S. (2013). *Tecnologia e Sociedade responsabilidades e desafios na Era da Informação Digital*. UFSCar, São Carlos, SP.
- [Ribeiro et al. 2020] Ribeiro, K. S. F. M., Maciel, C., e S. Bim, M. A. (2020). Gênero e tecnologias. In Maciel, C. e Viterbo, J., (eds), *Computação e sociedade: a profissão - volume 1*, pp. 104–140. EdUFMT Digital, Mato Grosso, Brasil.
- [Shilton 2018] Shilton, K. (2018). Values and ethics in human-computer interaction. *Foundations and Trends® in Human-Computer Interaction*, 12(2):107–171.
- [Silva 2020] Silva, T. (2020). *Comunidades, algoritmos e ativismos digitais: Olhares afrodiaspóricos*. LiteraRUA, São Paulo, SP.
- [Simon 2019] Simon, H. (2019). *The Sciences of the Artificial*. MIT Press, 3ª edição.

- [Spinello 2020] Spinello, R. A. (2020). *Cyberethics: Morality and Law in Cyberspace: Morality and Law in Cyberspace*. Jones & Bartlett Publishers, Burlington, MA, 7ª edição.
- [Stahl 2012] Stahl, B. C. (2012). Morality, ethics, and reflection: A categorization of normative is research. *JAIS*, 13(8):636–656.
- [Stair e Reynolds 2018] Stair, R. M. e Reynolds, G. W. (2018). *Principles of Information Systems*. CENGAGE Learning, Boston, MA, 13ª edição.
- [Tavani 2002] Tavani, H. (2002). The uniqueness debate in computer ethics: What exactly is at issue, and why does it matter? *Ethics and Information Technology*, 4.
- [Voltaire 2012] Voltaire, J. M. (2012). *Oeuvres Completes de Voltaire*, volume 12. Nabu Press.
- [Đurić e Gasevic 2013] Đurić, Z. e Gasevic, D. (2013). A source code similarity system for plagiarism detection. *The Computer Journal*, 56:70–86.

Capítulo

7

Desvendando a Camada de Aplicação na Internet das Coisas: Teoria, Prática e Tendências

Vagner E. Quincozes, Silvio E. Quincozes e Juliano F. Kazienko

Abstract

The Internet of Things (IoT) is a reality that is already part of our daily lives. In this context, it is essential to understand the fundamentals of prominent protocols, especially at the application layer level, as well as aspects for implementing solutions for IoT. In this document, we introduce the application layer and its protocols, namely MQTT, CoAP, MQTT-SN, XMPP and DDS. In addition, MQTT and CoAP protocols are presented in practice, as both have been gaining prominence in the literature and the industry, representing two alternatives for the IoT application layer. Finally, potential applications and trends in areas, namely: security, machine learning, computational paradigms, cyber-physical systems and user interfaces are addressed in order to provide insights to readers and encourage them to the development of IoT applications.

Resumo

A Internet das Coisas, do inglês, Internet of Things (IoT) é uma realidade que já faz parte do nosso cotidiano. Nesse contexto, é fundamental a compreensão dos fundamentos de protocolos proeminentes, especialmente a nível de camada de aplicação, bem como aspectos para a implementação de soluções na IoT. Neste documento, é apresentada a camada de aplicação e seus protocolos MQTT, CoAP, MQTT-SN, XMPP e DDS. Além disso, os protocolos MQTT e CoAP serão apresentados de maneira prática, visto que ambos vem ganhando destaque tanto na literatura quanto na indústria, representando duas alternativas para a camada de aplicação da IoT. Por fim, as potenciais aplicações e tendências nas áreas de segurança, aprendizado máquina, paradigmas computacionais para IoT, sistemas ciber-físicos e interfaces do usuário serão abordadas.

7.1. Introdução

Avanços em áreas como sensoriamento, sistemas embarcados e microeletrônica proporcionaram o surgimento da Internet das Coisas, do inglês, *Internet of Things* (IoT). Ela pode ser compreendida como uma extensão da Internet tradicional, que inicialmente era composta de sistemas finais preponderantemente computadores de mesa, oportunizando

a conexão de objetos variados de nosso dia-a-dia a grande rede [De Farias et al. 2019] [Santos et al. 2016] [Al-Fuqaha et al. 2015] [Tanenbaum and Wetherall 2011].

A IoT provocou uma quebra de paradigma na forma como entendemos redes de computadores, tornando mais atual o conceito de uma rede formada por dispositivos [Kurose and Ross 2010]. A ideia de um mundo interligado através de objetos inteligentes, isto é, com poder de processamento, comunicação e sensoriamento, proporciona benefícios, porém apresenta desafios de ordem técnica e social [Santos et al. 2016]. Dentre os desafios destacam-se, por exemplo, os recursos computacionais dos dispositivos, que muitas vezes são limitados, tornando-se necessária a adoção de novas tecnologias ou mesmo adaptação daquelas já existentes [Quincozes et al. 2021b].

Nesse contexto, os protocolos da camada de aplicação são responsáveis por efetuar a comunicação entre tais dispositivos, fornecendo interoperabilidade na interação humano-máquina e na comunicação entre máquinas [Borgiani et al. 2021]. Existem diversos protocolos disponíveis na camada de aplicação da IoT, a saber: *Advanced Message Queuing Protocol (AMQP)*, *Hyper Text Transfer Protocol Secure (HTTP)*, *Message Queuing Telemetry Transport (MQTT)*, *MQTT Sensor Network (MQTT-SN)*, *Data Distribution Service (DDS)*, *Extensible Messaging and Presence Protocol (XMPP)* e *Constrained Application Protocol (CoAP)*. Dentre eles, aqueles considerados mais proeminentes e com maior adoção pela indústria são o CoAP e o MQTT [Cosmi and Mota 2019] [Quincozes et al. 2019]. De forma geral, tais protocolos destacam-se pela sua adequação a aplicações que envolvam dispositivos com recursos restritos.

Como resultado de um processo de amadurecimento de tais protocolos da camada de aplicação, tendências e cenários de pesquisa têm despontado atualmente. Dentre eles, pode-se citar o estudo de técnicas de aprendizado de máquina a fim de agregar eficiência e segurança à operação desses protocolos. Nas páginas que seguem, essa e outras tendências proeminentes são discutidas [Tahsien et al. 2020] [Aazam et al. 2018] [Humayed et al. 2017].

Neste minicurso, a camada de aplicação na IoT e seus protocolos de comunicação são apresentados. Como prática, são demonstrados casos particulares de instalação, implementação e uso envolvendo os protocolos mais promissores: o MQTT e o CoAP. Adicionalmente, os potenciais rumos em termos de pesquisa e desenvolvimento relacionados aos protocolos dessa camada são apresentados. As principais contribuições deste trabalho são:

- (i) Apresentar a camada de aplicação da IoT e seus protocolos;
- (ii) Demonstrar casos práticos de como instalar, implementar e utilizar, em particular, os protocolos MQTT e CoAP;
- (iii) Discutir as potenciais aplicações e tendências de estudo envolvendo os protocolos da camada de aplicação relacionados à segurança da informação, aprendizado de máquina, os sistemas ciber-físicos, os paradigmas computacionais para IoT e suas interfaces do usuário.

O restante deste trabalho está organizado como segue. A Seção 7.2 aborda os fundamentos da IoT, com enfoque em sua camada de aplicação. Na Seção 7.3, os protocolos

da camada de aplicação são explanados. A Seção 7.4 apresenta os passos para instalação, implementação e utilização de casos particulares envolvendo os protocolos MQTT e CoAP. A Seção 7.5 discute as tendências relacionadas à temática tratada. Por fim, a Seção 7.6 apresenta a conclusão.

7.2. Fundamentos de IoT e a Camada de Aplicação

A camada de aplicação da IoT é composta por um conjunto de protocolos que comunicam dados através de *softwares* aplicativos, os quais estão distribuídos sendo executados em nós da rede. Nesta seção, são introduzidos conceitos importantes relacionados à IoT, com foco em sua camada de aplicação, a saber: evolução histórica, arquitetura de rede, sistemas operacionais, dispositivos e as arquiteturas de aplicação de rede.

7.2.1. Evolução Histórica

A Figura 7.1 ilustra uma linha do tempo com os principais marcos das últimas décadas que contribuíram para o desenvolvimento da IoT e seus protocolos da camada de aplicação.

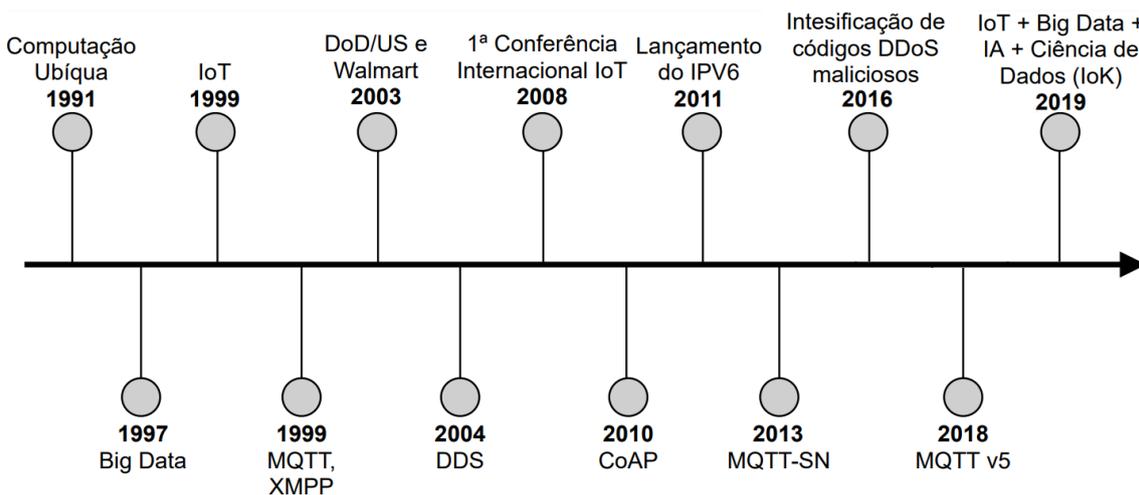


Figura 7.1. Linha do Tempo da IoT.

A ideia de conectar objetos começou a ser discutida na década de 1990. Mark Weiser trouxe o conceito de Computação Ubíqua em 1991 [Weiser 1991], com o objetivo de integrar a informática no cotidiano das pessoas, de forma em que os indivíduos não percebessem que estavam destinando comandos a computadores. Em 1997, surge o termo “*big data*”, que visa tratar (*i.e.*, processar e armazenar) grandes quantidades de dados [Cox and Ellsworth 1997]. Mais tardar, em 1999, Kevin Ashton cunhou o termo da Internet das Coisas (IoT), com a interconexão de objetos utilizados no dia a dia à Internet, surgindo os objetos “inteligentes” [Ashton et al. 2009].

Um marco importante a ressaltar consiste na publicação do protocolo MQTT, ocorrida no ano de 1999 [OASIS 2019]. Em meados de 2000, uma primeira versão do protocolo XMPP é disponibilizada publicamente [Saint-Andre et al. 2004]. Já no ano de 2003, instituições e empresas começam a utilizar de forma intensa um método de identificação através de sinais de rádio, denominada *Radio-Frequency IDentification*

(RFID). Em 2004, a versão 1.0 do protocolo DDS é publicada [Pardo-Castellote 2003]. Em 2008, aconteceu a primeira conferência internacional designada à IoT, realizada pela Zurich [Floerkemeier et al. 2008]. A partir deste momento, surgem novas discussões e debates sobre a IoT nos meios acadêmicos. Já no ano de 2010, surge o protocolo CoAP, especializado para dispositivos restritos [Shelby et al. 2014].

Com o intenso crescimento da IoT e de seus dispositivos, a disponibilidade dos endereços de Protocolos de Internet (IPs) estava quase esgotada. Então, em 2011, surge um novo Protocolo de Internet, denominado IPv6, aumentando a disponibilidade de IPs e colaborando ainda mais para o avanço da IoT. Em 2013, surge uma nova versão do protocolo MQTT, direcionada à redes de sensores (MQTT-SN) [Stanford-Clark and Truong 2013]. Em 2016, a IoT está em alta. Contudo, governos e empresas notam a intensificação de códigos maliciosos e de ataques de negação de serviço, do inglês *Distributed Denial of Service* (DDoS), percebendo a necessidade de combatê-los através da segurança da informação [Quincozes and Kazienko 2020]. Em 2018, a versão 5.0 do protocolo MQTT é disponibilizada. Por fim, em 2019 a discussão sobre a integração da IoT com Big Data, Inteligência Artificial e Ciência de Dados tem sido aprofundada.

7.2.2. Arquitetura IoT

Existem diversos protocolos que compõem a arquitetura da IoT, sendo que cada um deles opera em camadas específicas de acordo com os serviços ofertados em cada camada [Tanenbaum and Wetherall 2011]. Nesta seção, as camadas da arquitetura são apresentadas, destacando-se a camada de aplicação que é objeto de estudo neste documento.

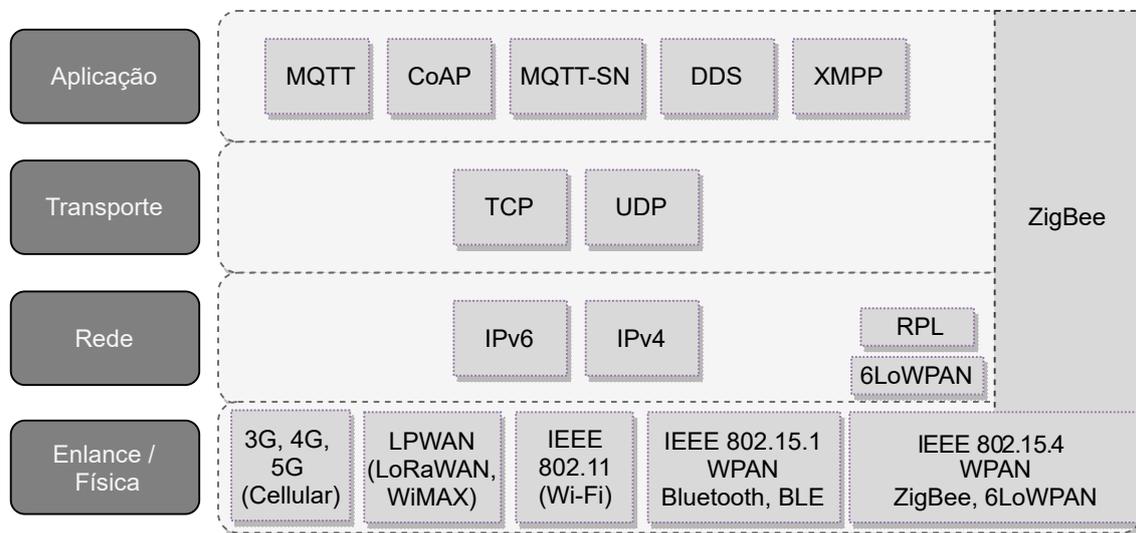


Figura 7.2. Arquitetura IoT. Adaptado de [Quincozes et al. 2019] e [Al-Fuqaha et al. 2015].

Conforme ilustrado na Figura 7.2, a primeira camada da arquitetura IoT é a de aplicação. Essa camada é composta por protocolos responsáveis pela comunicação entre dispositivos. Os protocolos MQTT [OASIS 2019], MQTT-SN [OASIS 2020], CoAP [Shelby et al. 2014], XMPP [Saint-Andre et al. 2004] e DDS [Pardo-Castellote 2003] são descritos e apresentados com mais detalhes na Seção 7.3.

A segunda camada da arquitetura IoT é a de transporte. A função dessa camada é realizar a transferência de dados entre máquinas. A transferência pode ser orientada a conexão ou sem conexão (datagrama). Existem dois principais protocolos nesta camada, que podem ser utilizados em aplicações IoT: o *Transmission Control Protocol* (TCP) [Postel 1981] e o *User Datagram Protocol* (UDP) [Postel 1980]. O protocolo de transporte TCP fornece confiabilidade na entrega de mensagens, pois garante que os dados transmitidos cheguem corretamente ao receptor. Já o protocolo UDP não é considerado confiável. Ele não controla o fluxo de transmissão e não retorna nenhuma mensagem confirmando que os dados foram entregues. Desse modo, não há garantia de que tais dados cheguem ao destino [Kurose and Ross 2010].

Os protocolos de Internet estão localizados na terceira camada e são responsáveis por endereçar, entregar e evitar engarrafamento na rede. Existem diferentes versões de protocolos de Internet, como o *Internet Protocol Version 4* (IPv4) e o *Internet Protocol Version 6* (IPv6). Atualmente, a versão mais recente é o IPv6, oficializado em junho de 2012. Há uma grande diferença entre tais protocolos no que diz respeito ao espaço de endereçamento. Ou seja, enquanto o IPv4 suporta 2^{32} , o IPv6 suporta 2^{128} endereços IPs. O IPv6 surgiu com o intuito de substituir o IPv4 gradualmente, visto que a capacidade de IPs disponíveis estava esgotada.

Na camada de rede, duas tecnologias se destacam. Primeiramente, *Routing Protocol for Low Power and Lossy Networks* (RPL) que é um protocolo de roteamento otimizado para redes com recursos computacionais restritos, atendendo requisitos de redes com baixa potência de dados e com perdas. O protocolo é baseado na teoria de grafos acíclicos direcionados formando um *Destination-Oriented Directed Acyclic Graph* (DODAG), onde os dados são dirigidos a um nó específico chamado raiz. A segunda tecnologia consiste no padrão *IPv6 over Low-Power Wireless Area Network* (6LoWPAN), que permite a comunicação de dados de forma eficiente em redes de baixa potência e velocidade. Ele atua adaptando o IPv6 para seu uso em redes que demandam baixo consumo IPv6 [Santos et al. 2016].

As últimas duas camadas da arquitetura, de enlace e física, podem ser definidas por várias tecnologias. Uma delas consiste no padrão IEEE 802.15.4, que permite controlar o acesso em redes sem fio pessoais com baixas taxas de transmissão [Quincozes et al. 2019] [Al-Fuqaha et al. 2015]. Outra tecnologia bastante adotada nesta camada é o *Long Range Wide Area Network* (LoRaWAN). Ela foi projetada para operar em redes de alto alcance e baixa potência, além de atender requisitos de baixo consumo energético. Na elaboração de uma rede, a tecnologia LoRaWAN implementa protocolos que definem a comunicação entre os componentes da arquitetura [Alliance 2017]. Adicionalmente, uma especificação importante para as camadas superiores consiste no ZigBee. O ZigBee tem como foco dispositivos de baixa potência e sua especificação prevê protocolos específicos executados nas camadas de rede e de aplicação [Alliance 2015].

7.2.3. Sistemas Operacionais e Dispositivos da IoT

Os sistemas operacionais (SO) da IoT são projetados para atender aos requisitos computacionais (capacidade de memória, energia, potência e armazenamento) dos dispositivos da IoT. Um dos sistemas operacionais mais utilizado por pessoas de todo o mundo é o Android, comumente adotado em *smartphones*. Também, ressalta-se a existência de

sistemas operacionais específicos para redes de sensores sem fio (RIOT) e microcontroladores (NuttX). Outros SO que serão apresentados consistem no Raspbian, TinyOS e Contiki, todos otimizados para operar em dispositivos com recursos restritos e atender os requisitos da IoT [Bansal and Kumar 2020][Santos et al. 2016][Hahm et al. 2015].

Android¹. É um SO baseado no núcleo Linux, revelado em 2007 pela Google junto com a fundação *Open Handset Alliance*, com o objetivo de desenvolver a indústria de dispositivos móveis [Alliance 2007]. O Android é o SO mais utilizado na atualidade. Esse SO pode ser instalado em *smartphones*, por exemplo, permitindo que os usuários controlem outros equipamentos da IoT. O *smartphone* funciona como um controle remoto [Santos et al. 2016].

RIOT. É um sistema operacional de código aberto, desenvolvido por uma comunidade de pessoas distribuídas pelo mundo todo². O RIOT fornece suporte à maioria dos dispositivos IoT, considerando baixa potência, microcontroladores (8, 16 e 32 bits) e demais dispositivos externos. As características do RIOT consistem no fácil desenvolvimento, suporte à segurança da camada de transporte (*e.g.*, DTLS), qualidade (*i.e.*, testes constantes) e com suporte às tecnologias da IoT, como os protocolos 6LoWPAN, MQTT-SN e CoAP³. Ademais, a RIOT conta com um ecossistema ciber-físico para monitorar e controlar objetos inteligentes [Bansal and Kumar 2020].

NuttX⁴. O NuttX é um sistema operacional que opera em tempo real. Ele pode ser construído como um microkernel ou como uma versão monolítica. É modular e escalável a ambientes de microcontroladores de 8 a 64 bits. Recentemente o NuttX começou a suportar a 6LoWPAN, tornando-se uma ótima opção para operar em dispositivos com recursos mais sofisticados, como os da IoT [Hahm et al. 2015].

Raspbian⁵. É um SO gratuito otimizado para o hardware Raspberry Pi, baseado no Debian. Vale ressaltar que o SO vem com mais de 35.000 pacotes e softwares pré-compilados para facilitar a instalação no Raspberry Pi. O Raspbian suporta autenticação, autorização e criptografia para multimídia [Bansal and Kumar 2020].

Contiki⁶. O Contiki é um sistema operacional de código aberto, sob a licença BSD-3-Clause⁷, desenvolvido para sistemas em rede com restrições de memória e foco em dispositivos da IoT sem fio de baixo consumo. A comunicação se dá através de protocolos padrão, como IPv6 / 6LoWPAN, 6TiSCH, RPL e CoAP [Hahm et al. 2015]. O Contiki é utilizado em vários projetos comerciais e não comerciais, como na iluminação pública, monitoramento de som para cidades inteligentes, medidores de energia elétrica em rede, dentre outras⁸. O Contiki fornece através do ContikiSec uma camada que provê serviços de segurança: autenticação, sigilo e integridade [Bansal and Kumar 2020].

¹Disponível em: <https://github.com/android>

²Disponível em: <https://github.com/RIOT-OS/RIOT>

³Disponível em: <https://www.riot-os.org/>

⁴Disponível em: <https://nuttx.apache.org/>

⁵Disponível em: <https://www.raspbian.org/>

⁶Disponível em: <https://github.com/contiki-os>

⁷Disponível em: <https://opensource.org/licenses/BSD-3-Clause>

⁸Disponível em: <https://github.com/contiki-os/contiki>

TinyOS⁹. Do mesmo modo que o Contiki, o TinyOS é um SO de código aberto licenciado pelo BSD e projetado para dispositivos sem fio de baixa potência, como os utilizados em redes de sensores, computação ubíqua, redes de área pessoal, edifícios inteligentes e medidores inteligentes [TinyOS 2013]. Além disso, o TinyOS possui suporte a biblioteca TinySec que fornece autenticação de mensagem, integridade e segurança na semântica de confidencialidade [Bansal and Kumar 2020]. Os microcontroladores do primeiro satélite da Estônia (ESTCube-1) utilizaram o TinyOS no módulo de comunicação do satélite [Sünter et al. 2016].

Ademais, sabe-se que os dispositivos e as aplicações da IoT crescem diariamente. Para citar alguns exemplos, destacamos os *smartphones*—compostos por sensores—e as lâmpadas inteligentes ou *smart lights*, que fornecem eficiência energética, conveniência e segurança. Há também aplicações IoT voltadas para o transporte, como os carros inteligentes, ou *smart cars*, que possuem autonomia para se movimentar, e os tratores inteligentes, ou *smart tractors*, que são capazes de realizar tarefas pré-programadas sem a necessidade de um operador. Outras aplicações IoT consistem em casas/edifícios inteligentes, ou *smart homes/buildings*, que possuem sistemas avançados para, por exemplo, controlar a temperatura [De Farias et al. 2019]. Mais dispositivos inteligentes são ilustrados na Figura 7.3.



Figura 7.3. Dispositivos Inteligentes que compõem a IoT.

Observa-se que a maior parte dos dispositivos possuem recursos computacionais limitados. Desse modo, a comunicação com tais dispositivos deve obedecer baixo consumo de energia e memória. Dentre as tecnologias que apresentam esses requisitos, destacamos o *Near Field Communication* (NFC), *Wireless Sensor Networks* (WSN), *Zigbee* e *Bluetooth*, que são adotados com frequência pela comunidade [Bansal and Kumar 2020].

7.2.4. Arquiteturas de Aplicação de Rede

Segundo [Kurose and Ross 2010], uma arquitetura de aplicação de rede determina a forma como a aplicação é organizada nos sistemas finais, além de ser projetada pelo desenvolvedor da aplicação. É diferente de arquitetura de rede, que é algo engessado do ponto de vista do desenvolvedor, fornecendo um conjunto específico de serviços às aplicações.

Essas arquiteturas definem a forma de comunicação entre as partes envolvidas. As arquiteturas de aplicação de rede mais adotadas na camada de aplicação da IoT são

⁹Disponível em: <https://github.com/tinyos>

[Kurose and Ross 2010]: *client/server*, exibido na Figura 7.4(a); e *peer-to-peer*, mostrado na Figura 7.4(b);

Outra arquitetura de aplicação a qual pode ser considerada como tal considerando o conceito acima é a *publish/subscribe*, ilustrada na Figura 7.4(c). Isso porque ela determina a forma como a aplicação está organizada nos sistemas finais: o *broker* sempre ativo e com maiores recursos computacionais, enquanto que o publicador e o assinante permanecem com a conexão intermitente. A seguir, tais arquiteturas são comentadas:

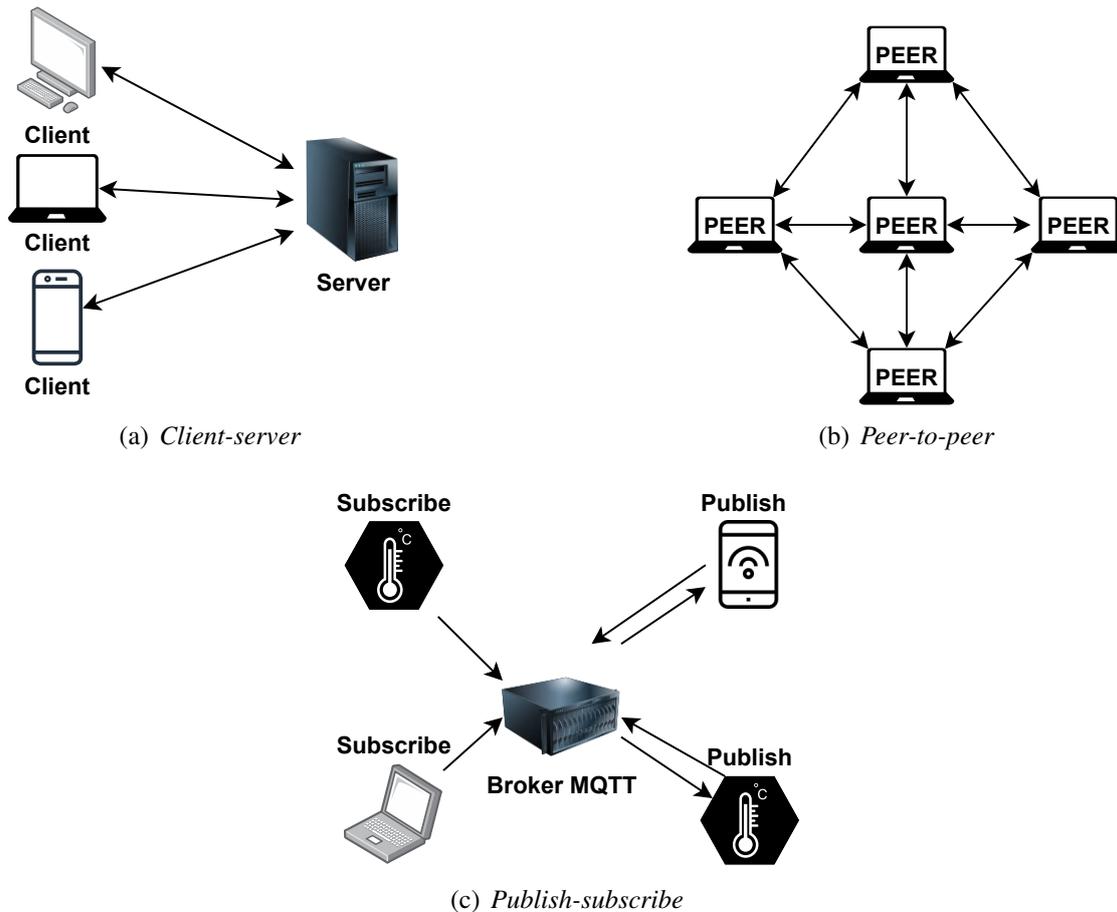


Figura 7.4. As Arquiteturas de Aplicação de Rede mais utilizadas na Camada de Aplicação da IoT.

- **Client-server.** Em português, é denominado como cliente-servidor. Essa arquitetura consiste em uma aplicação distribuída que compartilha tarefas e cargas de trabalho entre os provedores de recursos (*i.e.*, servidores) e os demandantes dos serviços (*i.e.*, clientes). Em outras palavras, os clientes solicitam informações ao servidor, enquanto que o servidor responde ao cliente compartilhando recursos. Essa estrutura costuma adotar o paradigma *request/response* na comunicação, que denota a demanda e a resposta por um serviço [Tanenbaum and Wetherall 2011].
- **Peer-to-peer.** Esse paradigma também é conhecido como par-a-par ou ponto-a-ponto (P2P). Nele, cada um dos dispositivos que compõem a rede funcionam tanto como cliente quanto como servidor. Desse modo, é possível compartilhar dados e serviços

sem a necessidade de um servidor central. Um exemplo de aplicação P2P pura consiste no Gnutella, que é uma rede de compartilhamento de arquivos em que as buscas são repassadas de um nó para o outro, dispensando um servidor centralizado. Ainda, é importante destacar a possibilidade de arquitetura híbrida formada a partir das arquiteturas cliente-servidor e par-a-par. O *torrent* é um exemplo de aplicação que utiliza arquitetura híbrida. Por exemplo, no *torrent* vários usuários podem realizar a conexão e transmitir arquivos, sem a necessidade de uma fonte intermediária para baixar os arquivos [Tanenbaum and Wetherall 2011].

- **Publish-subscribe.** Também conhecido por publicação-assinatura, esse paradigma é composto por três componentes principais: um publicador, um assinante e um servidor centralizado (*broker*). Basicamente, o *broker* é composto por tópicos. Os dispositivos publicadores são responsáveis por alimentar os tópicos do *broker*. Já os assinantes, são dispositivos interessados nas informações contidas nos tópicos, que assinam os mesmos para receber tais informações. Vale ressaltar que um *broker* pode conter múltiplos tópicos. Cada um deles pode receber mensagens de diversos dispositivos publicadores, as quais são entregues para todos os dispositivos que assinam aquele tópico [Quincozes et al. 2019].

Os protocolos da camada de aplicação podem utilizar tais arquiteturas expostas acima. Desse modo, é de extrema importância entendê-las para auxiliar na tomada de decisões no momento da implementação de novas aplicações IoT. Por exemplo, o protocolo MQTT é baseado na arquitetura *publish-subscribe* [Nebbione and Calzarossa 2020].

7.3. Protocolos da Camada de Aplicação

Nesta seção, os protocolos da camada de aplicação da IoT são apresentados, com ênfase nos protocolos MQTT e CoAP, que são considerados mais proeminentes na literatura. Ambos são considerados leves e recomendados para atuar em aplicações com recursos restritos. Por essa razão, apresentam grande adoção por parte da indústria. A Tabela 7.1 resume os protocolos e as suas principais características [Nebbione and Calzarossa 2020].

7.3.1. Protocolo CoAP

É um protocolo inspirado no HTTP e especializado para dispositivos com recursos restritos, conforme especificado na RFC 7252 [Shelby et al. 2014]. O protocolo CoAP utiliza o método de requisição/resposta: um cliente faz uma solicitação para a URI e o servidor responde. Na camada de transporte o CoAP adota o protocolo UDP para transferir mensagens [Tariq et al. 2020], considerado mais leve que o TCP. O protocolo define quatro tipos de mensagens utilizadas para fornecer confiabilidade, as quais são apresentadas a seguir: *Confirmable*, *Non-confirmable*, *Acknowledgment* e *Reset*. A Figura 7.5 ilustra exemplos de mensagens CON e NON.

Confirmable (CON). As mensagens do tipo CON visam prover confidencialidade. Desse modo, elas requerem confirmação quando chegam ao seu destino. Ou seja, se nenhum pacote for perdido durante a transmissão, uma mensagem de confirmação do tipo *Acknowledgment* é retornada. Caso contrário, uma mensagem do tipo *Reset* é repassada ao destinatário [Shelby et al. 2014].

Non-confirmable (NON). As mensagens do tipo NON são consideradas não confiáveis, visto que não requerem confirmação. As mensagens NON são interessantes para

Protocolo	Característica	Segurança	Transporte	Paradigma
MQTT	Flexibilidade, simplicidade e confiabilidade na entrega de mensagens.	TLS	TCP	<i>Publish/Subscribe</i>
CoAP	Foco em dispositivos com recursos restritos, alternativa ao HTTP.	DTLS	UDP	<i>Request/Response</i> <i>Publish/Subscribe</i>
MQTT-SN	Otimizado para redes de sensores. Possui modo de hibernação.	TLS	UDP	<i>Publish/Subscribe</i>
XMPP	Transferência de mensagens instantâneas (e.g., bate-papos ou videochamadas).	TLS	TCP	<i>Publish/Subscribe</i> <i>Request/Response</i>
DDS	Escalabilidade, confiabilidade, comunicação em tempo real - P2P, desempenho.	TLS/DTLS	TCP/UDP	<i>Publish/Subscribe</i>
HTTP	Recursos úteis e difundido. Entretanto, apresenta alto consumo de energia.	SSL	TCP	<i>Request/Response</i>

Table 7.1. Principais Protocolos da Camada de Aplicação da IoT. Adaptado de [Esfahani et al. 2017] e [Nebbione and Calzarossa 2020].

aplicações que fazem leituras repetidas regularmente, como as medições em sensores, por exemplo. Uma mensagem do tipo *Reset* é retornada caso o destinatário não seja capaz de processar a mensagem [Shelby et al. 2014].

Acknowledgment (ACK). Uma mensagem do tipo ACK confirma o recebimento de uma mensagem CON específica. A mensagem deve ser identificada por um ID de mensagem (e.g., ACK (ID: 0xVEq51)) [Shelby et al. 2014].

Reset (RST). Uma mensagem do tipo RST indica que uma mensagem do tipo NON foi recebida, mas não pode ser processada corretamente. Normalmente, isso acontece quando um nó receptor é reinicializado e um estado necessário para interpretar a mensagem é esquecido. Da mesma forma que a mensagem do tipo ACK, a RST requer um ID de identificação da mensagem [Shelby et al. 2014].

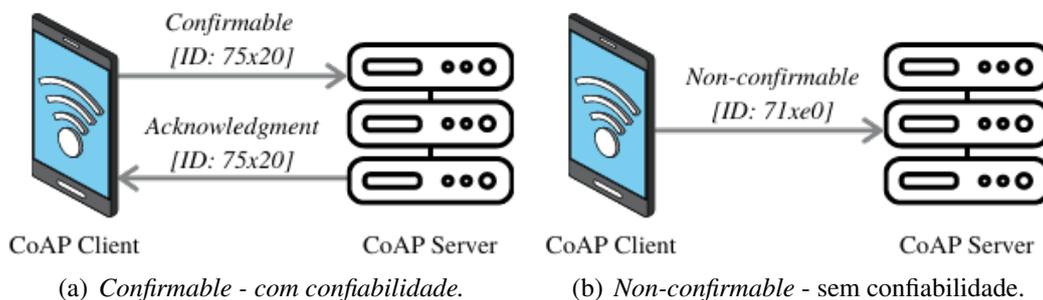


Figura 7.5. A obtenção de confiabilidade no protocolo CoAP. Adaptado de [Shelby et al. 2014].

7.3.2. Protocolo MQTT

O conceito de MQTT foi proposto por Andy Stanford-Clark e Arlen Nipper, em 1999. Vale ressaltar que em 2013, o MQTT tornou-se o protocolo padrão da organização para

o Avanço dos Padrões de Informação Estruturada (OASIS) [OASIS 2019]. Em síntese, o MQTT utiliza o paradigma de publicação/assinatura para comunicação, e foi projetado para suportar dispositivos com recursos restritos. Basicamente, a arquitetura do MQTT é composta por três tipos de componentes: (1) publicadores, (2) assinantes e um (3) *Broker*. Os dispositivos que desempenham papéis de assinantes são aqueles que se registram em tópicos de seu interesse no *broker*, de modo a receber mensagens transmitidas pelos dispositivos publicadores nos respectivos tópicos. Este paradigma é ilustrado na Figura 7.4(c). Diferente do protocolo CoAP, o MQTT utiliza o protocolo TCP na camada de transporte. Ademais, há três níveis de Qualidade de Serviço, ou em inglês, *Quality of Service* (QoS) do MQTT, que são responsáveis por entregar mensagens e garantir a precisão da comunicação: QoS-0, QoS-1 e QoS-2 [Quincozes et al. 2019].

QoS-0. Esse é o nível mais baixo de QoS do MQTT. Nele, não há garantia de que uma mensagem será entregue. Ademais, não é necessário que o destinatário confirme o recebimento da mesma. O QoS 0 é conhecido por “disparar e esquecer” a mensagem, visto que não fornece nenhum aviso ou garantia de entrega [OASIS 2019].

QoS-1. No nível de QoS 1, há garantia de que uma mensagem seja entregue pelo menos uma vez ao destinatário. Após disparar a mensagem, a mesma é armazenada pelo remetente até que uma mensagem de retorno do tipo *PUBACK* seja recebida, confirmando o recebimento. Nesse nível, a mensagem pode ser entregue uma ou múltiplas vezes. O pacote *PUBACK* contém um identificador que deve corresponder ao da mensagem *PUBLISH* [OASIS 2019].

QoS-2. O nível mais alto de QoS do MQTT é o 2. Nele, uma mensagem é entregue exatamente uma vez. Por isso, o QoS 2 é o mais seguro e, conseqüentemente, o mais lento. Existem pelo menos dois fluxos de solicitação para prover garantia de entrega: O remetente envia uma mensagem do tipo *PUBLISH* para o destinatário, que deve verificar a mensagem e responder ao remetente com um pacote *PUBREC*. Após obter a resposta, o remetente descarta a primeira mensagem *PUBLISH*, armazena o *PUBREC* recebido do destinatário e responde com um pacote *PUBREL*. A partir do momento que o destinatário obtém o pacote *PUBREL*, ele pode descartar todos os estados anteriores e responder o remetente com uma mensagem do tipo *PUBCOMP*. Da mesma maneira que o destinatário, quando o remetente recebe o *PUBCOMP*, ele pode descartar os estados anteriores. Todas as mensagens possuem um identificador do pacote [OASIS 2019].

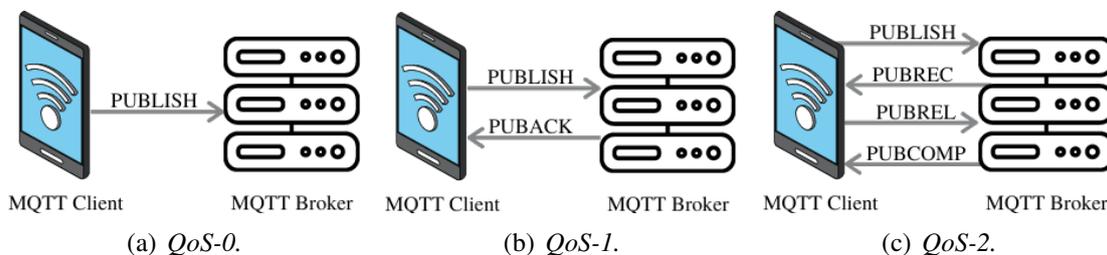


Figura 7.6. Os níveis de confiabilidade do protocolo MQTT. Essa é uma versão simplificada, visto que as mensagens de estabelecimento e fechamento de conexão estão omitidas. Adaptado de [OASIS 2019].

A Figura 7.6 ilustra a troca de mensagens em cada nível de confiabilidade do

MQTT. Por fim, vale ressaltar que o MQTT possui duas especificações: o MQTT v5.0 e o MQTT-SN, explanado a seguir.

7.3.3. Protocolo MQTT-SN

O MQTT-SN é uma versão adaptada do MQTT, específica para redes de sensores sem fio e com baixa largura de banda. Diferente do MQTT, o MQTT-SN utiliza o *User Datagram Protocol* (UDP) na camada de transporte para transferir mensagens. Uma das funcionalidades do MQTT-SN consiste na possibilidade de permanecer por determinado período de tempo em um modo chamado *sleeping*, onde o *broker* é responsável por armazenar todas as mensagens destinadas ao mesmo [Yassein et al. 2017] [Quincozes et al. 2019].

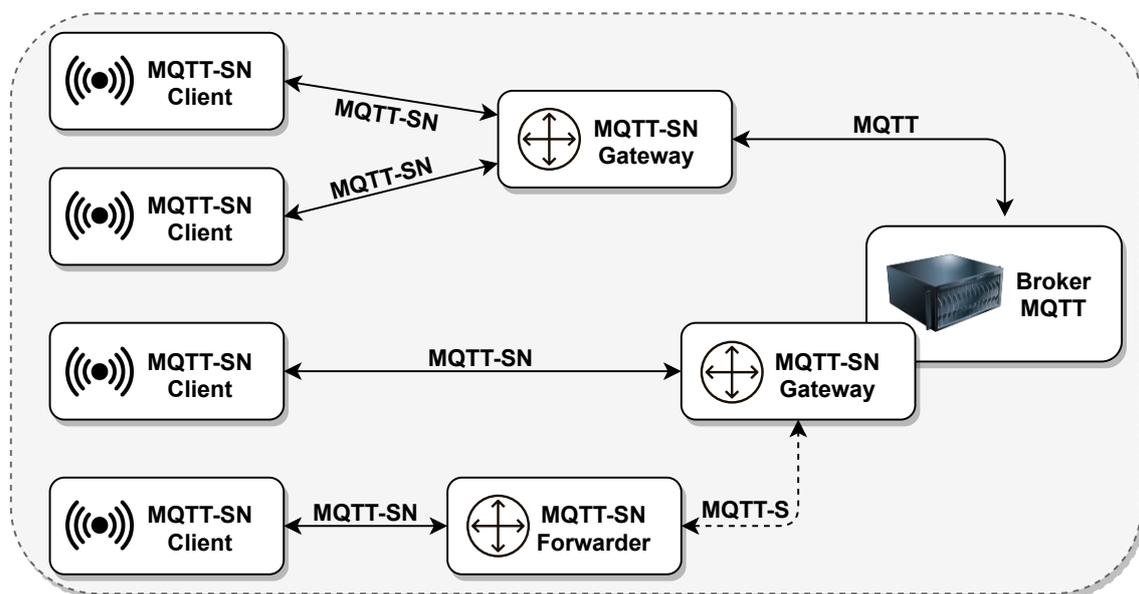


Figura 7.7. Arquitetura do MQTT-SN. Adaptado de [Stanford-Clark and Truong 2013].

A arquitetura do MQTT-SN é ilustrada na Figura 7.7. Tal arquitetura é composta por 3 tipos de componentes: MQTT-SN Client, MQTT-SN Gateway e MQTT-SN Forwarders. Normalmente, os Clients conectam-se ao Broker MQTT por meio dos Gateways utilizando o protocolo MQTT-SN. Vários Clients podem iniciar a conexão com um Gateway específico. Em alguns casos, o Client pode acessar o Broker através de Forwarders. Os Forwarders são responsáveis por encapsular os pacotes MQTT-SN e enviar para o Gateway [Stanford-Clark and Truong 2013].

7.3.4. Extensible Messaging and Presence Protocol (XMPP)

O XMPP é um protocolo aberto, extensível e adaptável, que surgiu em 1999 e teve sua primeira versão publicada em 2000. O XMPP se baseia no protocolo *Extensible Markup Language* (XML) para comunicação e transferência de arquivos em redes distribuídas. É um protocolo projetado para aplicações que exigem comunicação e mensagens instantâneas, como bate-papo em grupo ou vídeo-chamada e áudio. Além disso, possui uma funcionalidade que indica o status atual do cliente (*e.g.* online, offline ou ocupado)

[Çorak et al. 2018]. Nos últimos anos, o XMPP mostrou-se adequado para a IoT, reconquistando a atenção de muitos. O XMPP é executado sobre o TCP, suporta comunicação entre cliente-servidor e cliente-cliente, além disso, pode ser assíncrono utilizando o paradigma de publicação/assinatura ou síncrono através do modelo de solicitação/resposta. Por ser um protocolo projetado para aplicações quase em tempo real, o XMPP suporta mensagens de tamanho reduzido e de baixa latência [Karagiannis et al. 2015]. Empresas de grande porte utilizam o XMPP nas suas aplicações (*e.g. Google Talk, Whatsapp, Fortnite e League of Legends*).

7.3.5. Data Distribution Service (DDS)

A especificação do protocolo DDS foi aprovada pelo *Object Management Group* (OMG) em 2004, resultando na publicação da sua primeira versão. O DDS é um protocolo projetado para utilizar o paradigma de publicação/assinatura para comunicações *machine to machine* (M2M), que visa transmitir dados de forma escalável e em tempo real. Ao contrário do protocolo MQTT, o DDS adota a arquitetura *brokerless*, ou seja, sem a existência de nó intermediário. Ademais, ele utiliza *multicast* para fornecer comunicação confiável, eficiente e transmissão de informações para múltiplos destinatários simultaneamente, adaptando-se bem às aplicações IoT e M2M. No DDS, existe um conjunto com 23 políticas de QoS, que proporcionam a implementação de uma variedade de critérios nas comunicações, tais como: segurança, urgência, prioridade, durabilidade, propriedade, orçamento de latência, entre outras [Al-Fuqaha et al. 2015].

7.3.6. HTTP

Por fim, vale ressaltar a existência do protocolo HTTP, um dos principais protocolos utilizados em Sistemas de Informação para a transferência de dados utilizando o estilo arquitetural *REpresentational State Transfer* (REST) [Phung et al. 2020]. Embora seja um protocolo maduro, ele não é otimizado para IoT. Por exemplo, o CoAP difere do REST por adotar o protocolo UDP na camada de transporte, tornando-o mais apropriado para aplicações que utilizem dispositivos com recursos restritos na IoT. No entanto, o REST ainda é adotado por diversas plataformas devido a recursos úteis como negociação de tipo de conteúdo, mecanismos de autenticação e *caching* [Karagiannis et al. 2015].

A primeira versão do protocolo HTTP foi publicada na década de 1990, sob coordenação da *World Wide Web Consortium* (WWWC) e da *Internet Engineering Task Force* (IETF). O protocolo HTTP é comumente utilizado na Internet e é a base para efetivar a comunicação de dados da *World Wide Web* (WWW). O HTTP utiliza o paradigma de requisição-resposta e a comunicação cliente-servidor [Santos et al. 2016]. São definidos oito métodos que indicam a ação que deve ser realizada pelo servidor em uma requisição: *GET, HEAD, POST, PUT, DELETE, TRACE, OPTIONS e CONNECT*. Vale ressaltar que ao menos os métodos *GET* e *HEAD* devem ser implementados em um servidor HTTP para operação [Kurose and Ross 2010].

7.4. Prática

Nesta seção, a parte prática do minicurso é apresentada. Será dada ênfase na demonstração e configuração de implementações dos protocolos MQTT e CoAP, onde são disponibilizados exemplos de código-fonte com implementações de clientes utilizando bibliotecas

Java, como HiveMQ (para o protocolo MQTT) e CoAP Blaster (para o protocolo CoAP). Tais bibliotecas foram utilizadas para a implementação de um aplicativo *Android*, que é executado em um *smartphone*. Ademais, utilizou-se o *broker* MQTT Mosquitto para uma instalação local, em um dispositivo Raspberry Pi, que recebe mensagens publicadas pelo aplicativo instalado no *smartphone*. Por fim, um *laptop* foi utilizado para o desenvolvimento do aplicativo Android e monitoração de uso de recursos.

7.4.1. Implementação de Protótipo de clientes MQTT e CoAP em Android

De modo a demonstrar de maneira prática a implementação de dispositivos clientes CoAP e MQTT, foi implementado um protótipo através de um aplicativo Android que suporta ambos os protocolos. Tal protótipo é ilustrado na Figura 7.8, o qual é disponibilizado publicamente pelos autores deste manuscrito¹⁰. Para suportar os protocolos estudados, implementou-se um cliente publicador MQTT por meio da biblioteca disponibilizada em java HiveMQ (`com.hivemq-mqtt-client`) [HiveMQ 2012] e um cliente CoAP através da biblioteca CoAP Blaster (`com.google.iot.coap`) [CoapBlaster 2018].

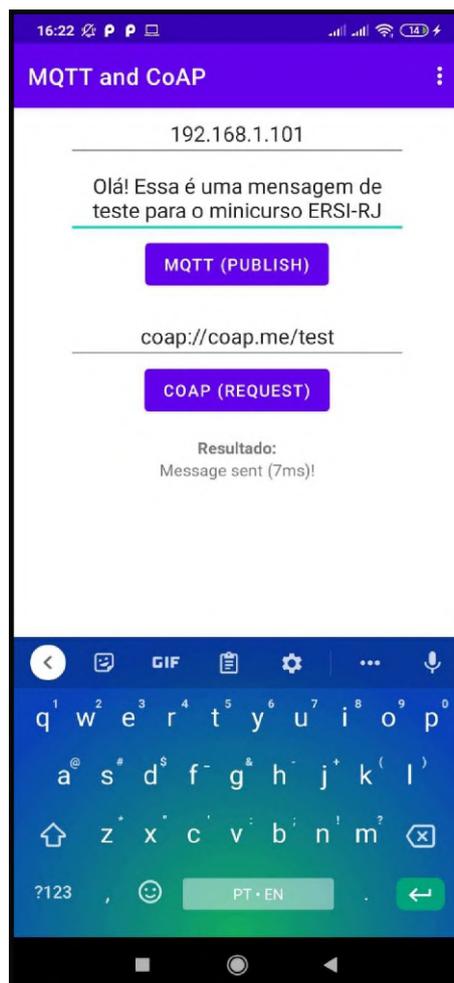
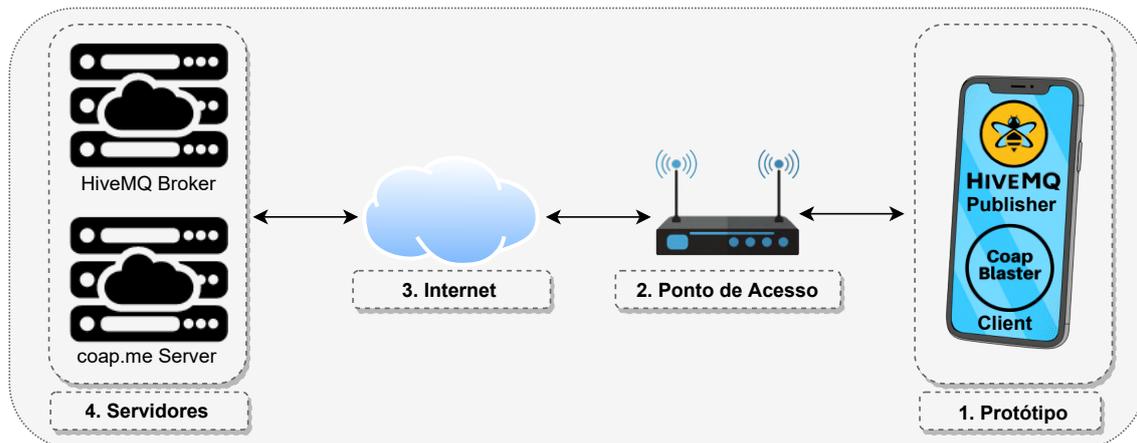


Figura 7.8. Implementação de protótipo de app Android para testes.

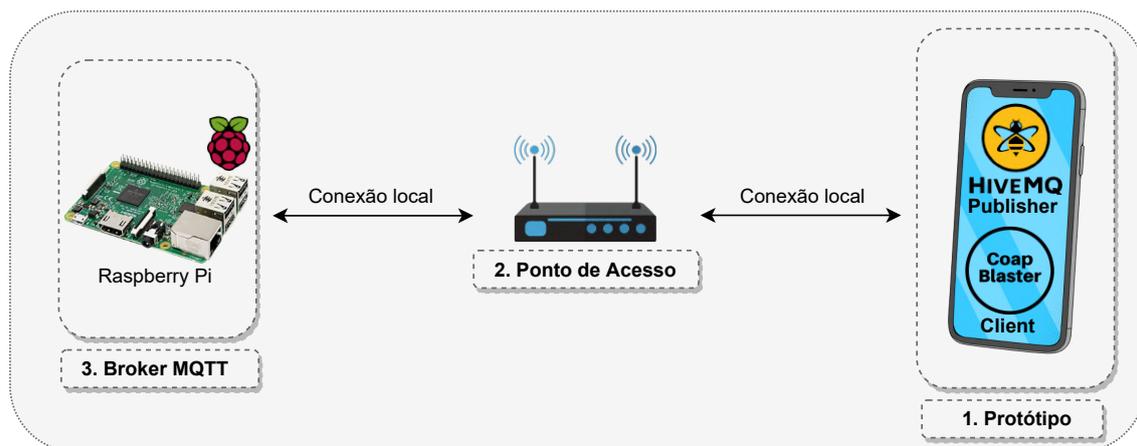
¹⁰Disponível em: <https://github.com/sequincozes/ERSI-RJ.git>

Tal protótipo inclui uma única tela com campos para o usuário informar o endereço do servidor CoAP ou broker MQTT. Desse modo, através dos botões MQTT (Publish) e CoAP (Request), os códigos que implementam a publicação de uma mensagem MQTT (listado no Código-fonte 7.1) e envio de uma requisição CoAP (listado no Código-fonte 7.2), respectivamente, são acionados.

Para fins de validação do protótipo, executaram-se testes, conforme a Figura 7.9(a), através da comunicação do *smartphone* contendo o aplicativo instalado com um *broker* MQTT e de um servidor CoAP publicamente disponíveis. Em síntese, a Figura 7.9(a) ilustra o primeiro cenário, no qual o protótipo (1) se associa ao ponto de acesso sem fio (2), conectando-se à Internet (3) a fim de se comunicar com o *broker* MQTT e o CoAP Server (4) públicos e disponíveis. Adicionalmente, a Figura 7.9(b) ilustra um cenário de validação em rede local. Desse modo, o protótipo (1) desenvolvido comunica através de um ponto de acesso sem fio (2) com *broker* MQTT, o qual é instalado em um dispositivo Raspberry Pi (3). O cenário local é discutido na Seção 7.4.2.



(a) Cenário 1: Broker MQTT e Servidor CoAP acessados em servidores publicamente acessíveis.



(b) Cenário 2: Broker MQTT instalado em um Raspberry Pi.

Figura 7.9. Cenários da implementação prática.

O código-fonte 7.1 ilustra a implementação prática de um cliente MQTT na linguagem de programação Java. As importações dos pacotes necessários são feitas nas

linhas 1 a 6. A classe `MqttClient` é definida na linha 8 e se estende até a linha 27. Nas linhas 11 a 17, o construtor da classe `MqttClient` é estabelecido. Também, existe um método responsável por publicar mensagens (linhas 19 a 26).

Perceba que a linha 1 do código-fonte 7.1 importa a biblioteca `MqttQos`, que é utilizada nas linhas 21 e 22 para definir o nível de confiabilidade da mensagem. Em síntese, três níveis de confiabilidade poderiam ser adotados, conforme visto anteriormente na Figura 7.6: `AT_MOST_ONCE` que representa o QoS-0, `AT_LEAST_ONCE` que implementa o QoS-1 e `EXACTLY_ONCE`, que faz referência ao QoS-2.

Código-fonte 7.1: Implementação de um Cliente MQTT em Java.

```

1 import com.hivemq.client.mqtt.datatypes.MqttQos;
2 import com.hivemq.client.mqtt.mqtt3.Mqtt3BlockingClient;
3 import com.hivemq.client.mqtt.mqtt3.Mqtt3Client;
4 import java.io.IOException;
5 import java.util.UUID;
6 import java.util.concurrent.TimeoutException;
7
8 public class MqttClient {
9     Mqtt3BlockingClient client;
10
11     public MqttClient (String uri) {
12         client = Mqtt3Client.builder()
13             .identifier(UUID.randomUUID().toString())
14             .serverHost(uri)
15             .buildBlocking();
16         client.connect();
17     }
18
19     public String publishMessage(String msg) {
20         long beginFullTime = System.currentTimeMillis();
21         client.publishWith().topic("test/topic").qos(MqttQos.
22             AT_LEAST_ONCE) .payload(msg.getBytes()).send();
23         long fullTime = System.currentTimeMillis() - beginFullTime;
24         client.disconnect();
25         return "Message sent (" +fullTime + "ms)!";
26     }
27 }

```

O código-fonte 7.2 refere-se a implementação prática de um cliente CoAP na linguagem de programação Java. Nele, as importações de pacotes são definidas nas linhas 1 a 9. Ressalta-se a existência de importações para exibir informações ao usuário (por exemplo, `Toast`), importações para tratar exceções de erros (por exemplo, `IOException` e `TimeoutException`) e importações específicas do CoAP, como `Client` que refere-se ao cliente a ser instanciado e `Message` que é uma biblioteca que dá suporte ao envio e recepção de mensagens.

Ademais, a classe principal, denominada `CoAPClient` tem início na linha 11 e término na linha 38. Tal classe é composta pelo seu construtor (linhas 14 a 23) e pelo método responsável pelo envio de mensagens (linhas 24 a 38).

Código-fonte 7.2: Implementação de um Cliente CoAP em Java.

```

1 import android.widget.Toast;
2 import com.google.iot.coap.Client;
3 import com.google.iot.coap.HostLookupException;
4 import com.google.iot.coap.LocalEndpointManager;
5 import com.google.iot.coap.Message;
6 import com.google.iot.coap.RequestBuilder;
7 import com.google.iot.coap.UnsupportedSchemeException;
8 import java.io.IOException;
9 import java.util.concurrent.TimeoutException;
10
11 public class CoAPClient {
12     RequestBuilder requestBuilder;
13
14     public CoAPClient(String uri) {
15         LocalEndpointManager manager = new LocalEndpointManager();
16         try {
17             Client client = new Client(manager, uri);
18             RequestBuilder = client.newRequestBuilder().setConfirmable(true);
19         } catch (UnsupportedSchemeException e) {
20             e.printStackTrace();
21         }
22     }
23
24     public String sendRequest(){
25         Message response = null;
26         try {
27             long beginFullTime = System.currentTimeMillis();
28             response = requestBuilder.send().getResponse();
29             long fullTime = System.currentTimeMillis() - beginFullTime;
30             return "Response received (" + fullTime + "ms):
31                 " + response.getPayloadAsString();
32         } catch (InterruptedException | HostLookupException | IOException |
33             TimeoutException e) {
34             e.printStackTrace();
35         }
36         return response.getPayloadAsString();
37     }
38 }

```

A ferramenta *Android Profiler* (ilustrada na Figura 7.10), do Ambiente de Desenvolvimento Integrado, do inglês, *Integrated Development Environment (IDE) Android Studio* permite a análise do uso de recursos das aplicações Android. Dessa forma, ao executar cada protocolo (*i.e.*, CoAP e MQTT) é possível analisar o consumo de rede, CPU, memória e energia. Adicionalmente, através do método *System.currentMillisseconds()*, nativo da linguagem Java, o tempo de resposta pode ser medido programaticamente.

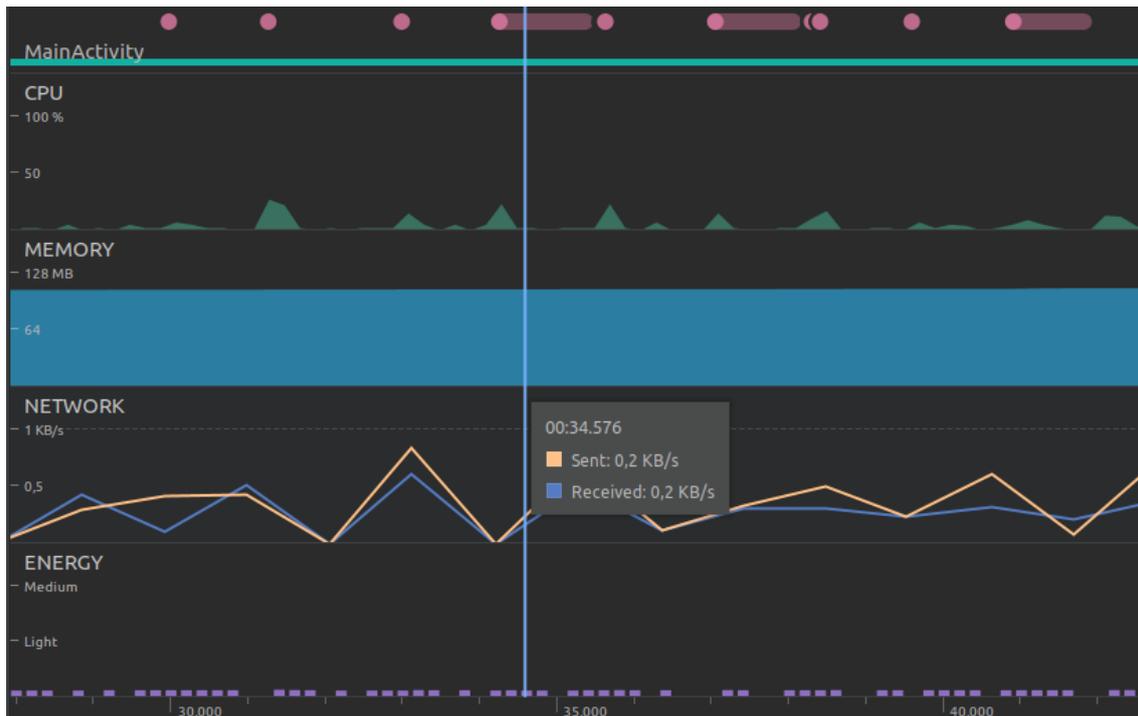


Figura 7.10. Android Profiler: Ferramenta de análise de uso de recursos do Android Studio.

7.4.2. Instalação e Demonstração de Uso de Broker MQTT em Raspberry Pi

A Figura 7.9(b) ilustra o cenário de rede local configurado para esta demonstração. De modo a instalar o *broker* Mosquitto no sistema operacional Raspbian, cujo é o sistema operacional padrão do Raspberry Pi, foram executados os seguintes comandos:

- Comando para atualização de pacotes e instalação do *broker* Mosquitto e seus clientes (publicador e assinante):

```
pi@raspberrypi:~ $ sudo apt update
pi@raspberrypi:~ $ sudo apt install -y mosquitto mosquitto-clients
```

- Comando para inicialização do serviço Mosquitto:

```
pi@raspberrypi:~ $ sudo systemctl enable mosquitto.service
```

- Comando para checar se o serviço está em execução:

```
pi@raspberrypi:~ $ mosquitto -v
```

- Comando para consultar o endereço IP do *broker* Mosquitto:

```
pi@raspberrypi:~ $ hostname -I
```

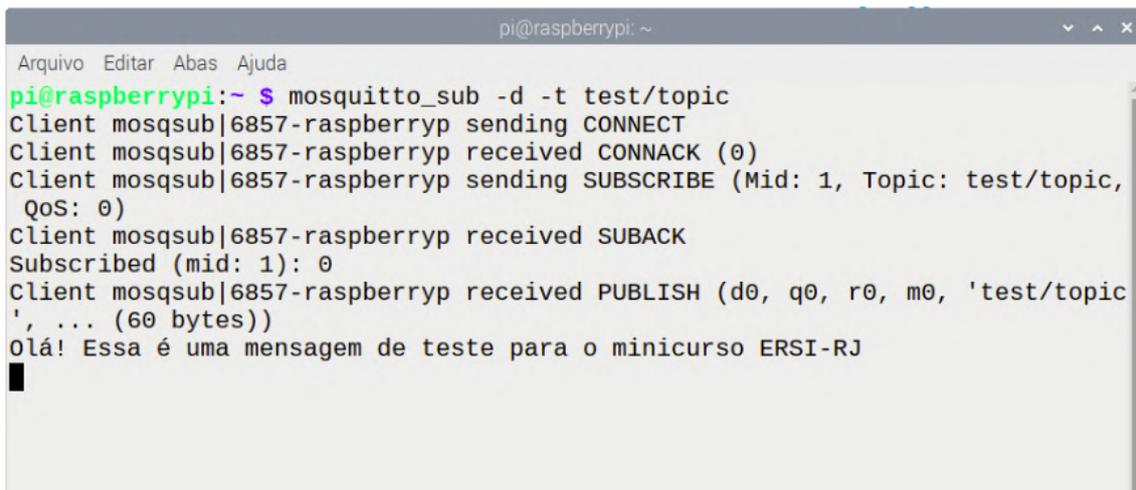
- Comando para executar um cliente assinante no Mosquitto:

```
pi@raspberrypi:~ $ mosquitto_sub -d -t testTopic
```

- Comando para executar um cliente publicador no Mosquitto, onde a mensagem `test` é enviada:

```
pi@raspberrypi:~ $ mosquitto_pub -d -t testTopic -m "test"
```

Note que através da instalação completa do Mosquitto é possível iniciar a execução de um *broker*, publicar mensagens e também assinar a tópicos. No entanto, optou-se pelo uso do protótipo apresentado na Seção 7.4.1 para a publicação de mensagens a partir do *smartphone*, através do protocolo MQTT. Na Figura 7.11, é ilustrado o dispositivo cliente que assina o tópico `test/topic` recebendo uma mensagem do tipo PUBLISH. É importante perceber que, para fins de demonstração, o assinante do referido tópico é executado na própria máquina onde roda o *broker*. Já a mensagem é publicada no mesmo tópico pelo dispositivo *smartphone*, usando o protótipo implementado.



```

pi@raspberrypi:~
Arquivo Editar Abas Ajuda
pi@raspberryp:~ $ mosquitto_sub -d -t test/topic
Client mosqsub|6857-raspberryp sending CONNECT
Client mosqsub|6857-raspberryp received CONNACK (0)
Client mosqsub|6857-raspberryp sending SUBSCRIBE (Mid: 1, Topic: test/topic,
QoS: 0)
Client mosqsub|6857-raspberryp received SUBACK
Subscribed (mid: 1): 0
Client mosqsub|6857-raspberryp received PUBLISH (d0, q0, r0, m0, 'test/topic
', ... (60 bytes))
Olá! Essa é uma mensagem de teste para o minicurso ERSI-RJ

```

Figura 7.11. Cliente assinante Mosquitto sendo executado no Raspberry Pi.

7.5. Tendências

Nesta seção, são apresentados assuntos que indicam tendências de estudo na IoT e, em particular, nos protocolos de sua camada de aplicação. Desse modo, são abordados tópicos, como a segurança, a utilização de técnicas de aprendizado de máquina, os sistemas ciber-físicos, os paradigmas computacionais e as interfaces do usuário.

7.5.1. Segurança da Informação

A segurança da informação é, cada vez mais, alvo de estudo em um mundo extremamente conectado. Ademais, o advento da IoT vem ampliando significativamente o número de dispositivos ligados à grande rede e impondo desafios no campo da segurança. Nos tempos atuais, intensificou-se a necessidade de resguardar um importante ativo de organizacional que consiste na informação.

Embora existam alternativas convencionais a fim de prover segurança à protocolos da camada de aplicação, como o *Transport Layer Security (TLS)* e o *Datagram TLS (DTLS)*, a adoção dessas opções pode não ser apropriada para dispositivos com recursos limitados, como muitos daqueles existentes no paradigma IoT. Isso tem oportunizado o surgimento de novas propostas envolvendo segurança na literatura especializada, conforme se descreve nos parágrafos seguintes.

Tendo em vista verificar a sobrecarga de mecanismos criptográficos simétricos na camada de aplicação da IoT, os autores de [Quincozes et al. 2021a] propõem estudo que envolve os protocolos CoAP e MQTT. O estudo é dirigido ao uso de cifras simétricas ao invés de cifras assimétricas, uma vez que, segundo os autores, aquelas possuem custo computacional e energético mais adequado para dispositivos com recursos limitado. Desse modo, algoritmos como o *Tiny Encryption Algorithm (TEA)*, *Data Encryption Standard (DES)* e *Advanced Encryption Standard (AES)* foram avaliados através de experimentos práticos disponibilizado publicamente¹¹. Os resultados demonstram que o mecanismo TEA é uma boa opção em termos de uso de memória, CPU, consumo energético, tempo de resposta e dados recebidos/enviados. No entanto, ressalta-se que algoritmos como o TEA e o DES estão na lista de cifras não recomendadas por órgãos internacionais, como o *National Institute of Standards and Technology (NIST)* [of Standards and Technology 2014]. Por outro lado, os resultados revelaram que a adoção do algoritmo AES, tanto para o protocolo CoAP quanto para o protocolo MQTT, implicou em maiores sobrecargas em termos de consumo energético e de tempos de resposta, entre outras métricas avaliadas. No entanto, o AES é uma cifra forte e recomendada pelo NIST.

De forma geral, ataques a sistemas de informação ferem a propriedades de segurança como a autenticidade, integridade, disponibilidade, confidencialidade, entre outras [Stallings 2015] [Kurose and Ross 2010] [Tanenbaum and Wetherall 2011]. Por exemplo, ataque de negação de serviço, do inglês, *Denial-of-Service (DoS)* afeta a disponibilidade dos sistemas. Já o ataque da interceptação afeta o sigilo das mensagens. Dessa forma, os trabalhos da literatura propõem mecanismos e arquitetura a fim de mitigar ataques em favor da garantia das propriedades citadas anteriormente.

Em [Rampelotto Junior et al. 2019], os autores propõem um mecanismo denominado *LegimateBroker* com o propósito de mitigar ataques de personificação em *broker* MQTT. A personificação consiste em ataque onde nó malicioso toma o lugar de nó legítimo, geralmente apropriando-se de identidade alheia. A proposta está calcada na autenticação mútua entre Publicadores e *Broker* do MQTT, armazenamento indireto de chaves no *Broker* e renovação periódica de chaves no *Broker* e no Publicador. Experimentos realizados indicam menor sobrecarga do *LegimateBroker* em termos de tempo médio para publicação e consumo energético quando comparado com outras abordagens, por exemplo, o uso de *Transport Layer Security (TLS)*.

O trabalho de [Khalil et al. 2020] apresenta um mecanismo que utiliza o protocolo CoAP para a descoberta de recursos (sensores, veículos, atuadores, *smartphones*, etc.) melhorando a autenticação, autorização e controle de acesso a tais recursos. O trabalho considera métricas, como o consumo de CPU, latência, consumo energético. Os experimentos utilizam como recursos um sensor de temperatura e uma tomada elétrica

¹¹Disponível em: <https://github.com/sequincozes/ExperimentationAPP2020>.

inteligente. A avaliação de desempenho realizada revelou baixa sobrecarga para todas as métricas retromencionadas quando comparada versão do CoAP que adota o mecanismo de segurança proposto com versão do CoAP que não adota qualquer mecanismo de segurança.

A fim de resistir a ataques como reprodução, *man-in-the-middle*, personificação e modificação, [Esfahani et al. 2017] apresentam um mecanismo baseado em operações ou-exclusivo e resumos criptográficos para autenticação mútua de dispositivos, entre outras propriedades de segurança, em redes IoT. Os autores defendem, através de sua análise de segurança, que o mecanismo provoca baixa sobrecarga em termos de custo computacional e de comunicação e, portanto é apropriado para dispositivos com recursos limitados usados na IoT, como sensores. Também, argumentam que sua proposta se aplica a comunicações *Machine-to-Machine* (M2M) em ambiente de IoT industrial. Para tanto, introduzem os protocolos da camada de aplicação da IoT—MQTT, AMQP, CoAP, XMPP, DDS—como protocolos M2M. A proposta prevê ainda a adoção de *Trusted Platform Module* (TPM) para proteção de chaves criptográficas em dispositivos. Embora o uso de TPM tenda a encarecer os dispositivos, especialmente em redes com larga escala, os autores demonstram preocupação com um importante tópico da segurança da informação: a gerência de chaves. Ela envolve aspectos como a geração, distribuição, armazenamento, renovação e destruição de chaves criptográficas [Menezes et al. 1996].

7.5.2. Aprendizado de Máquina

O Aprendizado de Máquina, do inglês, *Machine Learning* (ML) é uma técnica da Inteligência Artificial (IA) que treina máquinas utilizando algoritmos e é capaz de alterar o seu comportamento considerando apenas a experiência própria, em vez de programá-los, necessitando de pouca intervenção humana [Russell and Norvig 2013]. Desse modo, o ML vem sendo uma alternativa promissora para detectar ataques maliciosos e identificar comportamentos anormais através da análise dos mesmos. Recentemente, o ML tem sido amplamente adotado para o aperfeiçoamento de sistemas IoT, especialmente no que diz respeito a segurança da informação, dentre outras áreas [Tahsien et al. 2020].

Um importante direcionamento de pesquisas na área de ML consiste na adoção de suas técnicas para detectar ataques a protocolos da camada de aplicação da Internet das coisas [Tahsien et al. 2020]. Em particular, os protocolos desta camada estão suscetíveis a diversos ataques, como os de negação de serviço (DoS). Esse ataque objetiva consumir todos os recursos disponíveis de um sistema (e.g., memória e processamento). Com a sobrecarga gerada pelo ataque, normalmente a execução dos sistemas é interrompida, tornando-os indisponíveis aos usuários. Na literatura, existem trabalhos que propõem soluções para proteger os protocolos contra esse tipo de ataque. Por exemplo, [Syed et al. 2020] propõem um *framework* baseado em ML para detecção de ataques DoS dirigidos a *broker* MQTT. Os autores testaram a eficiência de sua proposta em três *brokers* MQTT *open-source*: Eclipse Mosquitto, VerneMQ e EMQ. O conjunto de dados (*dataset*) é produzido pelos próprios autores, além de não ser publicizado. Ele é composto por dados relativos a 33 *features* específicas do protocolo MQTT. Os experimentos realizados revelam que a proposta reduz as taxas de falsos positivos.

Outra tendência importante que se percebe consiste na crescente proposta e criação de *datasets* especializados para o estudo de ML junto a protocolos da camada de

aplicação da IoT. Uma dessas propostas consiste no MQTTset [Vaccari et al. 2020]. Nela, os autores propõem um *dataset* com foco no MQTT, composto por diferentes dispositivos da IoT (*i.e.*, sensores de movimento, temperatura e umidade). Esse *dataset* contém 11.915.716 amostras coletadas, incluindo tráfegos malignos e legítimos, sendo disponibilizada publicamente¹². O *dataset* cobre ataques cibernéticos na rede MQTT, como o *flooding Denial of Service (DoS)*, *MQTT Publish flood*, *SlowITe*, *malformed data* e *brute force authentication*. O *dataset* proposto conta com 33 *features*, sendo que aproximadamente 90% delas são específicas do protocolo MQTT. Os autores avaliaram o *dataset* considerando alguns algoritmos conhecidos de ML, como o *Random Forest* e *Naïve Bayes (NV)*. Os resultados obtidos mostram que os algoritmos tem uma precisão de detecção entre 87% e 91%, exceto o NV, onde os resultados ficam entre 64%.

Outro *dataset* proposto na literatura consiste no MQTT-IoT-IDS2020 Dataset, disponibilizado publicamente¹³ [Hindy et al. 2021]. Esse *dataset* é gerado com base em uma arquitetura de rede MQTT simulada. A rede é composta por doze sensores, um corretor, uma câmera simulada e um invasor. O *dataset* cobre quatro tipos de ataques: *aggressive scan*, *UDP scan*, *sparta SSH brute-force* e *MQTT brute-force attack*. O *dataset* é composto por 44 *features*, dividido entre *features* baseadas em pacotes, unidirecional e bidirecional. Ainda, os autores realizam experimentos usando técnicas de ML. Para facilitar a reprodução, os experimentos são disponibilizados publicamente em um repositório no GitHub¹⁴. Os resultados demonstram que as *features* baseadas em fluxo são mais adequadas para detectar ataques benignos e baseados em MQTT. Um resultado interessante é que a precisão aumentou de 72% para *features* baseadas em pacotes, enquanto que para o fluxo bidirecional e unidirecional ficou em torno de 99%.

Também, existem propostas onde os *datasets* são construídos com acesso mais restrito e com intuito de validar soluções propostas. Por exemplo, em [Hussain et al. 2021] os autores propõem uma estrutura para detectar ataques maliciosos em ambientes associados à saúde e IoT. A ferramenta IoT-Flock foi utilizada para gerar *dataset*. Essa ferramenta envolve dispositivos IoT e é capaz de gerar tráfegos normais e de ataques. Além disso, ela é capaz de gerar os ataques mais recentes e específicos para os protocolos MQTT e CoAP: *MQTT DDoS*, *MQTT publish flood*, *brute force*, *SlowITE*, além de suportar os ataques *MQTT Authentication Bypass Attack*, *MQTT Packet Crafting Attack* e *CoAP Replay Attack*. O *dataset* contém 10 *features*, sendo que 50% delas são específicas do protocolo MQTT. Ainda, para avaliar o desempenho do *dataset*, os autores utilizaram seis algoritmos de classificação de ML. Dentre os algoritmos avaliados, o *Random Forest* obteve o melhor desempenho na detecção de tráfego normal e malicioso, apresentando 99% de precisão, de *recall*, de acurácia e para F1-Score.

7.5.3. Sistemas Ciber-Físicos e Redes Industriais

Em um Sistema Ciber-Físico, do inglês, *Cyber-Physical System (CPS)*, os objetos físicos são equipados com dispositivos que possuem recursos de sensoriamento, atuação, computação e comunicação. Tais dispositivos coletam dados da camada de percepção e os

¹²Disponível em: <https://www.kaggle.com/cnrieiit/mqttset>

¹³Disponível em: <https://ieee-dataport.org/open-access/mqtt-internet-things-intrusion-detection-dataset>

¹⁴Disponível em: https://github.com/AbertayMachineLearningGroup/MQTT_ML

transmitem para a camada de aplicação, onde os centros de controle tratam dos processos físicos. A camada de aplicação contém aplicações especializadas para tarefas como como monitoramento, análise de dados e aplicações de usuário final [Quincozes et al. 2021a]. A arquitetura CPS tradicional de três camadas é apresentada na Figura 7.12. Protocolos como o MQTT e CoAP podem ser empregados para a transmissão de informações através das camadas representadas nessa figura.

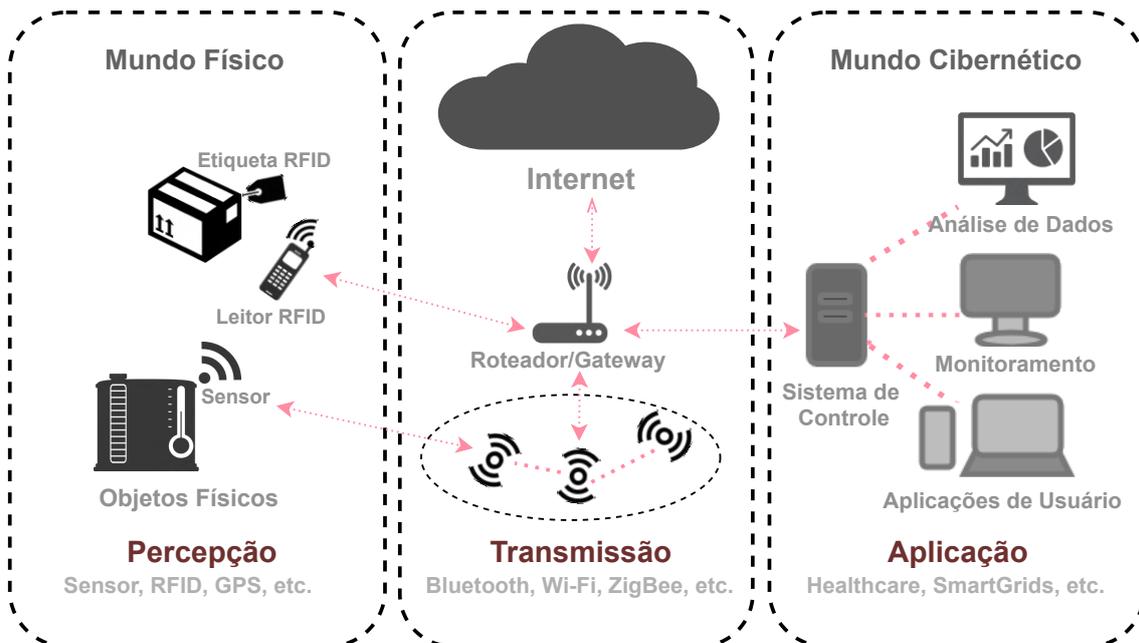


Figura 7.12. Arquitetura de CPSs (Adaptado de [Quincozes et al. 2021a] e [Santos et al. 2016]).

A camada de percepção é a principal novidade do CPS em relação aos sistemas de informação tradicionais. Ele contém dispositivos heterogêneos que coletam informações do mundo físico (*e.g.*, sensores e *tags* RFID), identificam certos eventos para responder e atuar de acordo com os mesmos. Esses dispositivos geralmente têm recursos limitados, incluindo a quantidade de memória, capacidade de processamento e fonte de energia [Quincozes et al. 2021a].

Os mundos físico e cibernético podem ser conectados por meio de tecnologias de rede com ou sem infraestrutura, como Wi-Fi, 3G / 4G / 5G, Bluetooth e Zigbee. Portanto, a camada de transmissão também envolve tecnologias e protocolos de comunicação heterogêneos. Esse é um dos principais motivos que leva a adoção de protocolos padronizados como o MQTT [Quincozes et al. 2021a].

Na camada de aplicação, é possível o controle, o monitoramento e o uso personalizado de dados do mundo físico no mundo cibernético. Existem várias aplicações CPS com características diferentes:

- **Smart Grids** adicionam monitoramento, controle e comunicação avançados à rede de energia elétrica. Eles permitem o uso eficiente de energia para consumidores, geradores e distribuidores [Quincozes et al. 2021a].

- **Veículos Inteligentes** comunicam-se entre si através do que pode ser chamado de CPS Veicular (VCPS). Os VCPSs podem ser utilizados para aumentar a eficiência energética e a capacidade rodoviária, por exemplo, usando uma metodologia baseada em pelotão [Jia et al. 2016], na qual os veículos em um grupo decidem cooperativamente sobre o comportamento de direção [Quincozes et al. 2021a].
- **Automação Industrial** é uma tendência que se tornou altamente prevalente e afeta os sistemas industriais modernos. A cooperação entre dispositivos distribuídos e software de controle torna os sistemas de produção mais adaptáveis, versáteis, escaláveis e responsivos. Com a capacidade de reagir em tempo real, os CPSs Industriais (ICPS) são fundamentais no conceito da Indústria 4.0 [Leitão et al. 2016].
- **Aplicações na Saúde**, onde encontram-se os Health-CPSs (HCPSs) lidam com grandes quantidades de dados complexos de saúde. Portanto, HCPSs devem garantir não apenas a interoperabilidade usando MQTT e CoAP, mas também devem atingir os requisitos de segurança, como privacidade, disponibilidade, autenticidade e integridade [Y. Zhang et al. 2017].

Os impactos dos sistemas ciber-físicos são visíveis na vida cotidiana, principalmente em áreas como elétrica, gás natural, distribuição de petróleo, eletrodomésticos, sistemas de saúde e transporte, dentre outros [Humayed et al. 2017]. Na indústria, os CPS permitem que as informações do espaço físico sejam monitoradas e sincronizadas com o espaço computacional cibernético. As análises avançadas das informações dão margem para as máquinas em rede trabalharem de forma colaborativa e eficiente, gerando uma transformação na indústria de manufatura para a Indústria 4.0 [Lee et al. 2015].

Grande parte dos sistemas ciber-físicos e redes industriais adotam protocolos da camada de aplicação da IoT, como o MQTT e CoAP, para prover interoperabilidade na comunicação. Tais adoções se dão pelas características dos mesmos: são protocolos abertos, simples e leves [Kirchhof et al. 2020]. Ademais, com a versão adaptada do MQTT, o MQTT-SN, existe a possibilidade de determinado dispositivo permanecer inativo por certo período de tempo [Quincozes et al. 2019]. Por isso, torna-se uma ótima opção para a economia de energia em redes industriais.

A Indústria 4.0 refere-se a quarta geração da indústria, que envolve a combinação de diversas tecnologias digitais e visa melhorar a produtividade de tecnologias de fabricação através da coleta e análise de dados em tempo real. Já é comum a projeção de plataformas para produtos e sistemas inteligentes com suporte aos protocolos da camada da aplicação da IoT [Aheleroff et al. 2020], como o MQTT. A Figura 7.13 ilustra a hierarquia entre a IoT, IIoT e Indústria 4.0 [Aazam et al. 2018].

7.5.4. Paradigmas computacionais

Os paradigmas computacionais em IoT são utilizados para entregar poder computacional sob demanda para diferentes partes da rede. Desse modo, paradigmas computacionais em IoT de Computação em Nuvem, Computação de Neblina e Computação de Borda, ou, em inglês, respectivamente, *Cloud Computing*, *Fog Computing* e *Edge Computing* são considerados [Mazon-Olivo and Pan 2021] [Prokhorenko and Babar 2020]. Tais paradigmas estão associados à IoT a fim de prover recursos, como poder de processamento e armazenamento, tendo em vista o grande volume de dados gerados pelos dispositivos de

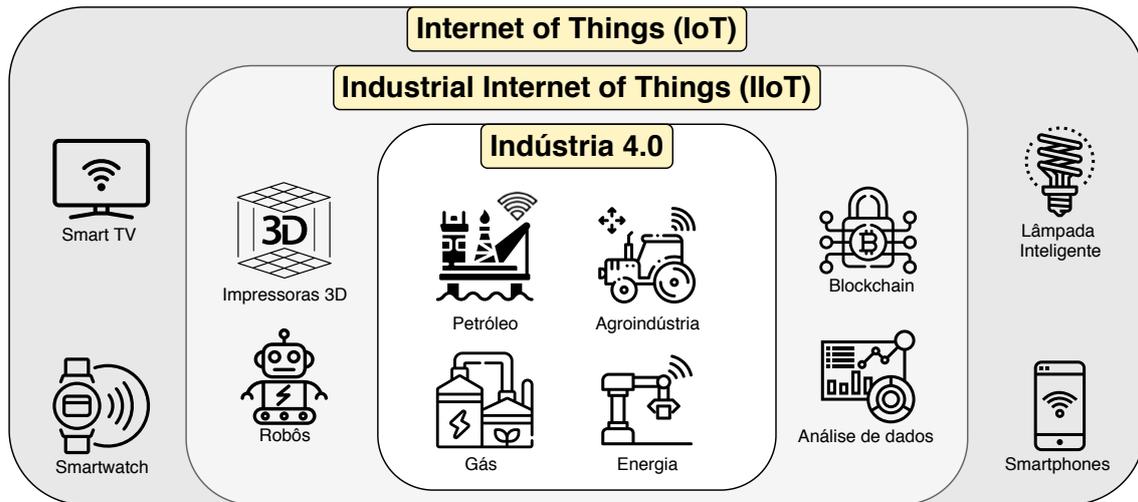


Figura 7.13. Hierarquia: IoT, IIoT e Indústria 4.0. Adaptado de [Aazam et al. 2018].

sensoriamento. Embora não sejam propriamente uma novidade, tais paradigmas são de extrema importância para as soluções que envolvam IoT [Mazon-Olivo and Pan 2021].

- **Cloud Computing.** Trata da disponibilização de serviços para empresas. Os principais serviços são processamento e armazenamento de dados em servidores, ou *datacenters*, que são acessíveis pela Internet. Esses serviços tendem a diminuir os custos e aumentar a flexibilidade das empresas [Prokhorenko and Babar 2020].
- **Fog Computing.** Trata-se da junção das soluções em nuvem e borda. O objetivo é promover uma arquitetura descentralizada, de modo em que as aplicações e o gerenciamento fossem distribuídos de forma inteligente entre a fonte de dados e a nuvem [Prokhorenko and Babar 2020]. Também, a *fog computing* vem sendo uma grande aposta para a Indústria 4.0 [Peralta et al. 2017] [Aazam et al. 2018].
- **Edge Computing.** Seu objetivo é semelhante ao da computação em nuvem. A principal diferença é que todo o processamento de dados acontece próximo dos dispositivos IoT, ou seja, na borda da rede onde eles se encontram. Dentre os principais benefícios, destacam-se a redução de banda e latência [Prokhorenko and Babar 2020].

A Figura 7.14 ilustra cada um dos paradigmas. Primeiramente, observa-se que a *Cloud Computing* é composta por milhares de *data centers*. As características deste paradigma consistem na centralização da lógica do negócio, suporte a operação com *big data*, visualização e análise de dados, alta capacidade de computação, dentre outras. A computação em nuvem opera sobre uma arquitetura centralizada. Em contraste com a *cloud computing*, o próximo paradigma, *fog computing*, é composto por milhões de *micro data centers*. Dentre as características deste paradigma, destacam-se a possibilidade de análise de dados, processamento e armazenamento temporal, resposta de controle e controle em tempo real. Este paradigma baseia-se em uma arquitetura distribuída de *fogs*. O último paradigma ilustrado na Figura 7.14, é o *edge computing*. Ele é composto por bilhões de dispositivos da IoT, como sensores e termostatos. Ressalta-se que a *edge computing* abrange sistemas embarcados, micro armazenamento, processamento em tempo real e análises de dados seguindo soluções da inteligência artificial e *machine-learning*.

Da mesma forma que a *fog computing*, esse paradigma opera com uma arquitetura distribuída [Mazon-Olivo and Pan 2021].

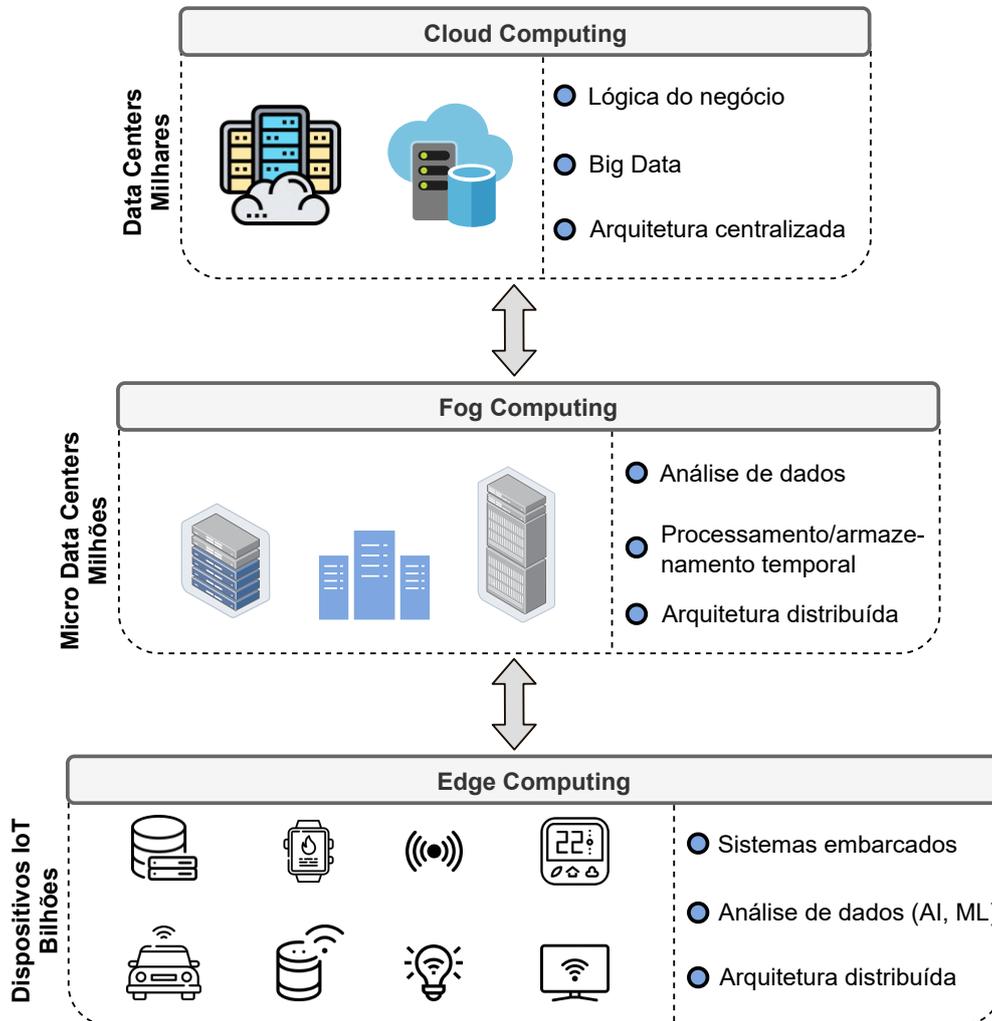


Figura 7.14. Os paradigmas computacionais *Cloud*, *Fog* e *Edge Computing*. Adaptado de [Mazon-Olivo and Pan 2021].

7.5.5. Interfaces do Usuário para IoT

Tais interfaces permitem gerenciar (*i.e.*, controlar, visualizar e analisar) dados de dispositivos da IoT em tempo real. A conexão dos dispositivos é feita por meio de protocolos da IoT, como o MQTT e o CoAP. Essas interfaces combinam características como escalabilidade, tolerância a falhas e desempenho. Existem inúmeras plataformas gratuitas e de código aberto, como *Bevywise*¹⁵, *ThingsBoard*¹⁶ e *IoTgo*¹⁷.

A *Bevywise* é uma plataforma que permite implantar aplicações IoT industriais e comerciais para coletar, analisar e visualizar dados históricos em tempo real. Ela disponibiliza uma API REST para facilitar a construção de aplicações móveis aos clientes. A

¹⁵Disponível em: <https://www.bevywise.com/iot-dashboard/>

¹⁶Disponível em: <https://thingsboard.io/>

¹⁷Disponível em: <https://github.com/itead/IoTgo>

plataforma implementa servidores e *brokers* MQTT e integra funcionalidades de aprendizado de máquina. A *Bevywise* apresenta soluções interessantes para a indústria 4.0, como o monitoramento e gerenciamento de produção, além de soluções futurísticas customizadas para alcançar maior produtividade¹⁸. Os principais produtos ofertados pela *Bevywise* são: *MQTT Broker*, *Hosted MQTT Broker*, *IoT Platform*, *IoT Simulator*, *IoT Dashboard*, *MQTT Gateway*, *L4Server - TCP/UDP* e *IoT Education Pack*¹⁹. A Figura 7.15 ilustra um exemplo do painel de monitoramento de produção em tempo real da plataforma.

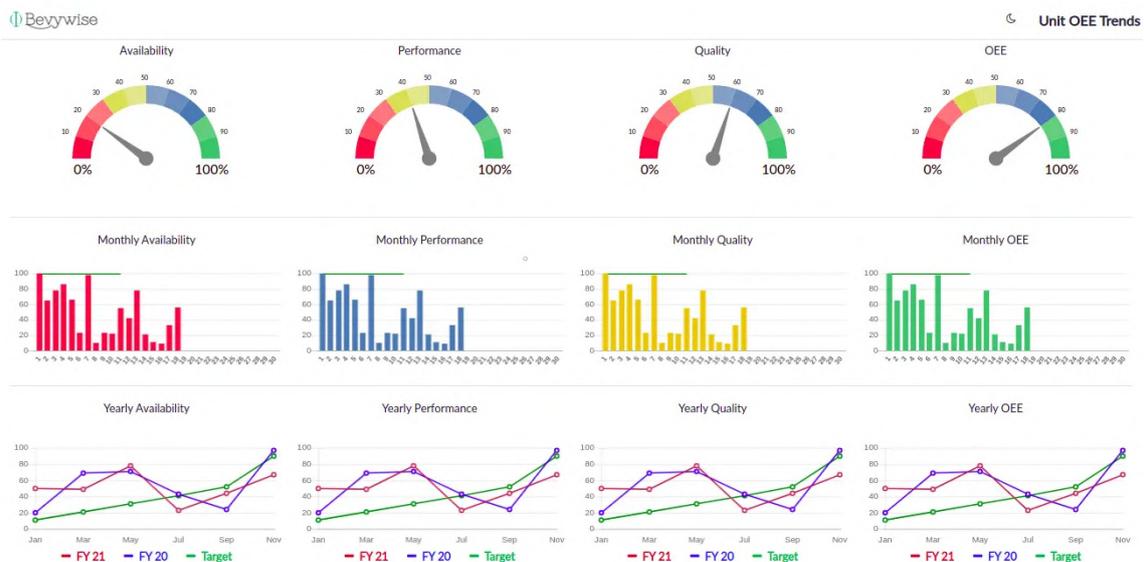


Figura 7.15. Painel de Monitoramento de Produção da Plataforma Bevywise. Retirado de [Bevywise 2021].

A *ThingsBoard* é uma plataforma IoT de código aberto para coleta de dados, processamento, visualização e gerenciamento de dispositivos. Os dispositivos adotam os protocolos da camada de aplicação da IoT, como o MQTT, CoAP e HTTP para prover conexão. A plataforma oferece suporte a implantações na nuvem e local, além de permitir a construção de *clusters* para obter o máximo de escalabilidade, tolerância a falhas e desempenho com a implementação da arquitetura de microsserviços. Outra característica dessa plataforma consiste na possibilidade de originar painéis para visualizar dados e controlar remotamente dispositivos da IoT em tempo real. Para isso, a plataforma disponibiliza mais de 30 *widgets* personalizáveis. A ferramenta fornece suporte para aplicações no contexto de energia inteligente (*i.e., smart energy*), agricultura inteligente (*i.e., smart farming*), medições inteligentes (*i.e., smart metering*) e rastreamento de frota (*i.e., fleet tracking*)²⁰. A Figura 7.16 ilustra o painel de monitoramento para agricultura inteligente.

Assim como *ThingsBoard*, a *IoTgo* é uma plataforma de código aberto que permite que qualquer pessoa implemente o seu próprio serviço de nuvem. Segundo os autores, a plataforma foi projetada para ser aberta, gratuita, simples e de fácil utilização.

¹⁸Disponível em: <https://www.bevywise.com/industry-4-0/solutions/>

¹⁹Disponível em: <https://www.bevywise.com/>

²⁰Disponível em: <https://github.com/thingsboard/thingsboard>



Figura 7.16. Painél de Controle Agrícola Inteligente da Plataforma ThingsBoard. Retirado de [ThingsBoard 2021].

A documentação oficial²¹ apresenta as informações necessárias para instalação, configuração e execução da IoTgo, além de uma biblioteca com a relação de dispositivos compatíveis. A plataforma fornece suporte às interfaces *mobile*, *web* e *desktop*. Além disso, ela disponibiliza uma API que, quando integrada às interfaces, permite que os clientes controlem os dispositivos. A plataforma fornece suporte aos protocolos HTTP e WebSocket. No entanto, os criadores recomendam a adoção do WebSocket devido às características: permite o envio da atualização do status do dispositivo tanto para o dispositivo real quanto para os clientes do proprietário do dispositivo.

7.6. Conclusão

Neste minicurso, os fundamentos da IoT foram apresentados, desde a fundamentação histórica, arquitetura, sistemas operacionais, dispositivos, arquiteturas de aplicação de rede e, com maior ênfase, a sua camada de aplicação. Protocolos dessa camada como XMPP, DDS e HTTP REST, CoAP e MQTT foram apresentados, sendo os dois últimos alvo de maior destaque. Realizou-se demonstração prática que consistiu na implementação de um protótipo de aplicativo em Android que instância clientes MQTT e CoAP através das bibliotecas HiveMQ [HiveMQ 2012] e CoapBlaster [CoapBlaster 2018]. O protótipo de aplicativo está disponibilizado publicamente pelos autores deste manuscrito²². A implementação contou com dois cenários: (i) Broker MQTT e Servidor CoAP acessados em servidores publicamente disponíveis na Internet; e (ii) Broker MQTT instalado na plataforma Raspberry Pi acessado através do protótipo em ambiente de rede local. No segundo cenário, demonstra-se os comandos para instalação e uso do Broker MQTT.

Adicionalmente, assuntos relevantes que indicam tendências de estudo na IoT e

²¹Disponível em: <https://github.com/itead/IoTgo>

²²Disponível em: <https://github.com/sequincozes/ERSI-RJ.git>

nos protocolos da camada de aplicação foram apresentados. O primeiro deles consiste na segurança da informação, onde foi indicada a clara necessidade de mecanismos customizados para rodar em dispositivos da IoT. Um segundo assunto trata-se do aprendizado de máquina, que vem sendo uma alternativa promissora para detectar ataques maliciosos nos protocolos da camada de aplicação da IoT. Particularmente, percebe-se forte tendência na construção de *datasets* especializados para o estudo de protocolos da camada de aplicação. O terceiro refere-se aos sistemas ciber-físicos e redes industriais, que estão cada vez mais presentes em aplicações como *smart grids*, veículos inteligentes, automação industrial e saúde. Outro assunto tratado consiste nos paradigmas computacionais *cloud*, *fog* e *edge computing*, utilizados para entregar poder computacional sob demanda para diferentes partes da rede. Nota-se, que é uma tendência crescente a integração de IoT com tais paradigmas computacionais a fim de melhorar, por exemplo, o tempo de resposta das aplicações, capacidade de armazenamento e processamento. Por fim, as interfaces do usuário para IoT foram apresentadas. Elas permitem controlar, visualizar e analisar dados de dispositivos e têm papel importante uma vez que apresentam os dados oriundos dos dispositivos IoT ao usuário, suportando o uso de protocolos da camada de aplicação.

Espera-se que este documento possa auxiliar no processo de tomada de decisão por parte de desenvolvedores e analistas na implantação de novas aplicações IoT. Para tanto, buscou-se trazer a luz os fundamentos de IoT, bem como os protocolos mais recomendados que operam em nível de camada de aplicação e suas características. Através da implementação prática, espera-se motivar novos usuários e desenvolvedores que estão entrando nesta área. Por fim, acredita-se que a discussão envolvendo assuntos e tendências possa nortear os próximos passos daqueles que buscam ingressar, ou até mesmo os já iniciados, na temática tratada neste minicurso.

Referências

- Aazam, M., Zeadally, S., and Harras, K. A. (2018). Deploying Fog Computing in Industrial Internet of Things and Industry 4.0. *IEEE Transactions on Industrial Informatics*, 14(10):4674–4682.
- Aheleroff, S., Xu, X., Lu, Y., Aristizabal, M., Velásquez, J. P., Joa, B., and Valencia, Y. (2020). Iot-enabled smart appliances under industry 4.0: A case study. *Advanced engineering informatics*, 43:101043.
- Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., and Ayyash, M. (2015). Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376.
- Alliance, L. (2017). LoRaWAN® Specification v1.1. Disponível em: https://loralliance.org/resource_hub/lorawan-specification-v1-1/. Acessado em: Agosto/2021.
- Alliance, O. H. (2007). Industry Leaders Announce Open Platform for Mobile Devices. Disponível em: http://www.openhandsetalliance.com/press_110507.html. Acessado em: Agosto/2021.
- Alliance, T. Z. (2015). ZigBee Pro Specification - Connectivity Standards Alliance. Disponível em: <https://zigbeealliance.org/wp-content/uploads/2019/11/docs-05-3474-21-0csg-zigbee-specification.pdf>. Acessado em: Agosto/2021.

- Ashton, K. et al. (2009). That ‘Internet of Things’ Thing. *RFID Journal*, 22(7):97–114.
- Bansal, S. and Kumar, D. (2020). Ecosystem: A Survey on Devices, Gateways, Operating Systems, Middleware and Communication. *International Journal of Wireless Information Networks*, 27:340–364.
- Bevywise (2021). Bevywise. Disponível em: <https://www.bevywise.com/industry-4-0/manufacturing-execution-system/>. Acessado em: Agosto/2021.
- Borgiani, V., Moratori, P., Kazienko, J. F., Tubino, E. R., and Quincozes, S. E. (2021). Toward a Distributed Approach for Detection and Mitigation of Denial-of-Service Attacks Within Industrial Internet of Things. *IEEE Internet of Things Journal*, 8(6):4569–4578.
- CoapBlaster (2018). CoapBlaster. Disponível em: <https://github.com/google/coapblaster>. Acessado em: Agosto/2021.
- Çorak, B. H., Okay, F. Y., Güzel, M., Murt, Ş., and Ozdemir, S. (2018). Comparative Analysis of IoT Communication Protocols. In *2018 International symposium on networks, computers and communications (ISNCC)*, pages 1–6. IEEE.
- Cosmi, A. B. and Mota, V. F. (2019). Uma Análise dos Protocolos de Comunicação para Internet das Coisas. In *III Workshop de Computação Urbana*, pages 153–166.
- Cox, M. and Ellsworth, D. (1997). Application-Controlled Demand Paging for Out-of-Core Visualization. In *Proceedings. Visualization’97 (Cat. No. 97CB36155)*, pages 235–244. IEEE.
- De Farias, C., Rodrigues Caldas de Aquino, G., Costa, G., Kopp, L. F., and Campos, B. (2019). Fusão de dados para Ambientes Inteligentes. In *Livro de Minicursos da VI Escola Regional de Sistemas de Informação do Rio de Janeiro (ERSI-RJ 2019)*, chapter 5, pages 133–157.
- Esfahani, A., Mantas, G., Maticsek, R., Saghezchi, F. B., Rodriguez, J., Bicaku, A., Maksuti, S., Tauber, M. G., Schmittner, C., and Bastos, J. (2017). A Lightweight Authentication Mechanism for M2M Communications in Industrial IoT Environment. *IEEE Internet of Things Journal*, 6(1):288–296.
- Floerkemeier, C., Langheinrich, M., Fleisch, E., Mattern, F., and Sarma, S. E. (2008). *The Internet of Things: First International Conference, IOT 2008, Zurich, Switzerland, March 26-28, 2008, Proceedings*, volume 4952. springer.
- Hahm, O., Baccelli, E., Petersen, H., and Tsiftes, N. (2015). Operating Systems for Low-End Devices in the Internet of Things: a Survey. *IEEE Internet of Things Journal*, 3(5):720–734.
- Hindy, H., Bayne, E., Bures, M., Atkinson, R., Tachtatzis, C., and Bellekens, X. (2021). Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset). In *Selected Papers from the 12th International Networking Conference*, pages 73–84. Springer, Springer International Publishing.
- HiveMQ (2012). HiveMQ. Disponível em: <https://www.hivemq.com/>. Acessado em: Agosto/2021.

- Humayed, A., Lin, J., Li, F., and Luo, B. (2017). Cyber-Physical Systems Security – A Survey. *IEEE Internet of Things Journal*, 4(6):1802–1831.
- Hussain, F., Abbas, S. G., Shah, G. A., Pires, I. M., Fayyaz, U. U., Shahzad, F., Garcia, N. M., and Zdravevski, E. (2021). A Framework for Malicious Traffic Detection in IoT Healthcare Environment. *Sensors*, 21(9):3025.
- Jia, D., Lu, K., Wang, J., Zhang, X., and Shen, X. (2016). A survey on platoon-based vehicular cyber-physical systems. *IEEE Commun. Surveys Tuts.*, 18(1):263–284.
- Karagiannis, V., Chatzimisios, P., Vazquez-Gallego, F., and Alonso-Zarate, J. (2015). A Survey on Application Layer Protocols for the Internet of Things. *Transaction on IoT and Cloud computing*, 3(1):11–17.
- Khalil, K., Elgazzar, K., Abdelgawad, A., and Bayoumi, M. (2020). A security approach for CoAP-based internet of things resource discovery. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pages 1–6.
- Kirchhof, J. C., Michael, J., Rumpe, B., Varga, S., and Wortmann, A. (2020). Model-driven Digital Twin Construction: Synthesizing the Integration of Cyber-Physical Systems with Their Information Systems. In *23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, pages 90–101.
- Kurose, J. F. and Ross, K. W. (2010). *Redes de Computadores e a Internet: Uma Abordagem Top-Down*. São Paulo: Addison Wesley, 5.ed.
- Lee, J., Bagheri, B., and Kao, H.-A. (2015). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing letters*, 3:18–23.
- Leitão, P., Colombo, A. W., and Karnouskos, S. (2016). Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges. *Computers in Industry*, 81:11–25.
- Mazon-Olivo, B. and Pan, A. (2021). Internet of Things: State-of-the-art, Computing Paradigms and Reference Architectures. *IEEE Latin America Transactions*, 100(1e).
- Menezes, A. J., Oorschot, P. C. V., and Vanstone, S. A. (1996). *Handbook of Applied Cryptography*. Boca Raton, FL: CRC Press.
- Nebbione, G. and Calzarossa, M. C. (2020). Security of IoT Application Layer Protocols: Challenges and Findings. *Future Internet*, 12(3).
- OASIS (2019). MQTT Version 5.0 OASIS Standard. Disponível em: <https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>. Acessado em: Agosto/2021.
- OASIS (2020). MQTT for Sensor Networks (MQTT-SN) Version 1.3. Disponível em: <https://www.oasis-open.org/committees/download.php/66972/mqtt-sn-v1.3-wd02.docx>. Acessado em: Agosto/2021.
- of Standards, N. I. and Technology (2014). FIPS 140-2 Security Policy. Disponível em: <https://csrc.nist.gov/csrc/media/projects/cryptographic-module-validation-program/documents/security-policies/140sp2092.pdf>. Acessado em: Agosto/2021.

- Pardo-Castellote, G. (2003). *OMG Data-Distribution Service: Architectural Overview*. In *23rd International Conference on Distributed Computing Systems Workshops, 2003. Proceedings.*, pages 200–206. IEEE.
- Peralta, G., Iglesias-Urkia, M., Barcelo, M., Gomez, R., Moran, A., and Bilbao, J. (2017). *Fog Computing Based Efficient IoT Scheme for the Industry 4.0*. In *2017 IEEE international workshop of electronics, control, measurement, signals and their application to mechatronics (ECMSM)*, pages 1–6. IEEE.
- Phung, C. V., Dizdarevic, J., and Jukan, A. (2020). *An Experimental Study of Network Coded REST HTTP in Dynamic IoT Systems*. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.
- Postel, J. (1980). *RFC 768: User Datagram Protocol*. Disponível em: <https://rfc-editor.org/rfc/rfc768.txt>. Acessado em: Agosto/2021.
- Postel, J. (1981). *RFC 793: Transmission Control Protocol*. Disponível em: <https://rfc-editor.org/rfc/rfc793.txt>. Acessado em: Agosto/2021.
- Prokhorenko, V. and Babar, M. A. (2020). *Architectural Resilience in Cloud, Fog and Edge Systems: A Survey*. *IEEE Access*, 8:28078–28095.
- Quincozes, S., Emilio, T., and Kazienko, J. F. (2019). *MQTT Protocol: Fundamentals, Tools and Future Directions*. *IEEE Latin America Transactions*, 17(09):1439–1448.
- Quincozes, S. E. and Kazienko, J. F. (2020). *Machine Learning Methods Assessment for Denial of Service Detection in Wireless Sensor Networks*. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pages 1–6. IEEE.
- Quincozes, S. E., Mossé, D., Passos, D., Albuquerque, C., Ochi, L. S., and dos Santos, V. F. (2021a). *On the Performance of GRASP-Based Feature Selection for CPS Intrusion Detection*. *IEEE Transactions on Network and Service Management*.
- Quincozes, V. E., Quincozes, S. E., and Kazienko, J. F. (2021b). *Avaliando a Sobrecarga de Mecanismos Criptográficos Simétricos na Internet das Coisas: Uma Comparação Quantitativa entre os Protocolos MQTT e CoAP*. In *XX Workshop em Desempenho de Sistemas Computacionais e de Comunicação*, pages 13–24. SBC.
- Rampelotto Junior, C., Quincozes, S. E., and Kazienko, J. F. (2019). *LegitimateBroker: Mitigando Ataques de Personificação em Broker MQTT na Internet das Coisas*. In *XIX Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, pages 141–154. SBC.
- Russell, S. J. and Norvig, P. (2013). *Inteligência artificial*. GEN LTC, 3.ed.
- Saint-Andre, P. et al. (2004). *Extensible Messaging and Presence Protocol (XMPP): Core*. Disponível em: <https://datatracker.ietf.org/doc/html/draft-ietf-xmpp-3920bis>. Acessado em: Agosto/2021.
- Santos, B. P., Silva, L. A., Celes, C. S., Borges Neto, J. B., Peres, B. S., Vieira, M. A. M., Vieira, L. F. M., Goussevskaia, O. N., and Loureiro, A. A. (2016). *Internet das Coisas: da Teoria à Prática*. In *XXXIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, chapter 1, pages 1–50.

- Shelby, Z., Klaus, H., and Bormann, C. (2014). The Constrained Application Protocol (CoAP). RFC 7252, Universitaet Bremen TZI. Acessado em: Agosto/2021.
- Stallings, W. (2015). *Criptografia e Segurança de Redes: Princípios e Práticas*. São Paulo: Pearson, 6.ed.
- Stanford-Clark, A. and Truong, H. L. (2013). MQTT for sensor networks (MQTT-SN) protocol specification. *International business machines (IBM) Corporation version*, 1(2).
- Sünter, I., Slavinskis, A., Kvell, U., Vahter, A., Kuuste, H., Noorma, M., Kutt, J., Vendt, R., Tarbe, K., Pajusalu, M., et al. (2016). Firmware Updating Systems for Nanosatellites. *IEEE Aerospace and Electronic Systems Magazine*, 31(5):36–44.
- Syed, N. F., Baig, Z., Ibrahim, A., and Valli, C. (2020). Denial of service attack detection through machine learning for the IoT. *Journal of Information and Telecommunication*, 4(4):482–503.
- Tahsien, S. M., Karimipour, H., and Spachos, P. (2020). Machine Learning Based Solutions for Security of Internet of Things (IoT): A Survey. *Journal of Network and Computer Applications*, 161:18.
- Tanenbaum, A. S. and Wetherall, D. (2011). *Computer networks, 5th Edition*. Prentice Hall Professional Technical Reference.
- Tariq, M. A., Khan, M., Raza Khan, M. T., and Kim, D. (2020). Enhancements and Challenges in CoAP—A Survey. *Sensors*, 20(21):6391.
- ThingsBoard (2021). ThingsBoard. Disponível em: <https://thingsboard.io/smart-farming/>. Acessado em: Agosto/2021.
- TinyOS (2013). TinyOS. Disponível em: <http://www.tinyos.net/>. Acessado em: Agosto/2021.
- Vaccari, I., Chiola, G., Aiello, M., Mongelli, M., and Cambiaso, E. (2020). MQTTset, a New Dataset for Machine Learning Techniques on MQTT. *Sensors*, 20(22):6578.
- Weiser, M. (1991). The Computer for the 21st Century. Disponível em: http://wiki.daimi.au.dk/pca/_files/weiser-orig.pdf. Acessado em: Agosto/2021.
- Y. Zhang et al. (2017). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1):88–95.
- Yassein, M. B., Shatnawi, M. Q., Aljwarneh, S., and Al-Hatmi, R. (2017). Internet of Things: Survey and open issues of MQTT Protocol. In *2017 International Conference on Engineering & MIS (ICEMIS)*, pages 1–6. IEEE.



Wagner Ereno Quincozes possui o título de Engenheiro de Software pela Universidade Federal do Pampa (UNIPAMPA) e atualmente é mestrando em Engenharia de Software (UNIPAMPA). Tem interesse nas linhas de pesquisas relacionadas a Segurança da Informação, Redes de Computadores, Provedores de Serviços de Internet, Engenharia de Software, Aprendizado de Máquina e Internet das Coisas. Atualmente participa do projeto de Inovação em Provedores de Serviço de Internet Regionais, no Laboratório de Estudos Avançados (LEA) da UNIPAMPA. Também, participa do projeto de pesquisa e extensão MARFIM - Métodos de Aprendizado de Máquina aplicados à

Segurança em Redes Formadas por Dispositivos com Recursos Limitados, na Universidade Federal de Santa Maria (UFSM).

CV Lattes: <http://lattes.cnpq.br/5834103446295383>



Silvio Ereno Quincozes é professor no IFRS (Canoas). Engenheiro de Software (UNIPAMPA, Brasil), Mestre em Ciência da Computação (UFSM, Brasil) e doutorando em Ciência da Computação (UFF, Brasil), com período sanduíche na University of Pittsburgh (Pitt, EUA). Seus principais interesses de linha de pesquisa incluem Segurança da Informação, Redes de Computadores, Internet das Coisas, Mineração de Dados e Engenharia de Software. Também têm experiências profissionais como Consultor de Sistemas, Desenvolvedor Android, Desenvolvedor PHP e Analista de Dados. Atualmente é bolsista pelo Programa de Excelência Acadêmica da CAPES (Proex). É

membro do projeto "Teleproteção em IEC-61850" e do Laboratório MidiaCom. Revisor dos periódicos IEEE Access e Computers Security, e dos simpósios Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT) e Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC).

CV Lattes: <http://lattes.cnpq.br/9401130360785458>



Juliano Fontoura Kazienko possui graduação nos cursos de Administração e de Ciência da Computação pela Universidade Regional Integrada do Alto Uruguai e Missões (URI) e Universidade do Vale do Itajaí (UNIVALI), respectivamente. Concluiu seu mestrado em Ciência da Computação pela Universidade Federal de Santa Catarina (UFSC). Concluiu o doutorado em Computação pela Universidade Federal Fluminense (UFF), com período sanduíche na Universidade da Califórnia em Irvine (UCI), Estados Unidos. Em 2013, recebeu premiação da Secretaria de Assuntos Estratégicos da Presidência da República (SAE/PR) em razão de sua tese de doutorado. Exerceu o cargo de professor adjunto na Universidade Federal do Pampa (UNIPAMPA) no período de 2013 até

2016. Atualmente, é professor adjunto da Universidade Federal de Santa Maria (UFSM). Tem experiência em ciência da computação, especialmente na área de segurança em computação. Seus temas de interesse são: Segurança da Informação, Redes Sem fio e Móveis, Redes Centradas em Informação, Inteligência Artificial aplicada a Redes e Internet das Coisas.

CV Lattes: <http://lattes.cnpq.br/7847000086448712>

Capítulo

8

A linguagem R na análise de dados: Um estudo de caso dos transportes públicos do RJ durante a pandemia da Covid-19

Julia Amaro Gonçalves Fagundes, Matheus Henrique de Sousa Oliveira e Vladimir Fagundes

Abstract

The use of the R Programming Language for statistical analysis using techniques of trend and seasonality reduction, Pearson correlation and cross correlation proved to be very relevant to solve real world problems. In this context, this chapter was developed in order to develop a tutorial applied to a relevant and current topic (the Covid-19 pandemic and public transport), aiming at a fast and targeted learning, so that the student quickly reaches his goals and ensure a solid foundation if you are interested in delving into both the subject and the language.

Resumo

A utilização da Linguagem de Programação R para elaboração de análise estatística utilizando técnicas de redução de tendência e sazonalidade, correlação de Pearson e correlação cruzada mostrou-se bastante relevante para resolver problemas do mundo real. Neste contexto, este capítulo foi desenvolvido com o intuito de apresentar um tutorial aplicado a um tema relevante e atual (a pandemia da COVID-19 e os transportes públicos), objetivando um aprendizado rápido e direcionado, de maneira que o aluno atinja rapidamente seus objetivos e garanta uma base sólida, caso possua interesse de se aprofundar tanto no assunto quanto na linguagem.

8.1. Introdução

Em março de 2020, à época com 118.326 casos confirmados e 4.292 óbitos, a Organização Mundial da Saúde (OMS) declarou a disseminação da Covid-19 como pandemia [OMS, 2020]. Desde então, o mundo enfrenta uma emergência de saúde pública, com poder devastador ainda não experimentado em um passado recente. O novo coronavírus é uma doença respiratória aguda e, por vezes, grave, podendo levar o infectado a óbito. De acordo com Perlan et al. (2020) a transmissão interpessoal ocorre pelo contato com secreções contaminadas, principalmente pelo contato com grandes gotículas respiratórias, mas também pode ocorrer por meio do contato com uma superfície contaminada.

Entretanto, apesar das características da transmissão de vírus, as medidas de limitação da circulação de pessoas para conter a propagação do Covid-19 não podem ser confundidas com a paralisação do transporte público. Apesar da redução do número de pessoas em circulação nas ruas, a interrupção dos serviços não é uma opção; pelo contrário. Nem todos os usuários de transporte público podem trabalhar remotamente ou possuem veículo próprio para se locomoverem quando necessário [Lima et al., 2020].

Diante desse cenário de incerteza, surge a necessidade de avaliar o comportamento da transmissão para que possam ser ofertadas alternativas para que profissionais dos serviços essenciais possam se locomover de maneira segura e adequada. Além disso, existe um grande número de dados espalhados em diversos órgãos que necessitam de análise para que se possa gerar indicadores que balizem tomadas de decisão com maior acurácia e possibilitem desenvolver políticas públicas que vão ao encontro dos anseios da sociedade.

Neste cenário, a associação de dados reais relativos à pandemia e aos transportes públicos permite uma análise mais assertiva em termos de estimativa da necessidade de ações preventivas, bem como para estratégias de retomada. Cabe ressaltar que uma maneira eficiente de se estimar a relação entre diferentes conjuntos de dados é a análise de correlação, onde são estudadas potenciais relações de causa e efeito dentre os dados disponíveis. Para tal estudo, podem ser utilizadas diferentes linguagens de programação como ferramentas para manipular e analisar dados tais como Python, Julia, Matlab, Java e R.

A linguagem de programação R se apresenta como uma alternativa robusta e eficiente para a ciência de dados em geral. Por ser uma linguagem desenvolvida por estatísticos, muitas funções e fórmulas que nas linguagens mais difundidas precisam ser desenvolvidas do zero, no R já vêm prontas para aplicação com breves comandos. Esta característica, além de permitir códigos mais simples, também contribui para minimização de erros. Cabe ainda ressaltar que o R é gratuito e de código aberto, o que gera potencial vantagem competitiva em relação a ferramentas como SAS e SPSS.

Dentre as principais funções da linguagem está extensa relação de modelos estatísticos, que vão desde a modelagem linear e não-linear, a análise de séries temporais, os testes estatísticos clássicos, análise de agrupamento e classificação, etc, além da apresentação gráfica dos resultados contando com variadas técnicas, passando

também pela criação e manipulação de mapas. Desta forma, a análise estatística da linguagem 'R' se apresenta como uma boa alternativa para análise de dados de forma simples e dinâmica, uma vez que dispõe de uma enorme quantidade de pacotes desenvolvidos com as principais manipulações mais comumente utilizadas na análise de dados, o que facilita tanto o aprendizado quanto a utilização da linguagem..

A escolha dos dados aqui aplicados foi feita a título de demonstração da utilização da linguagem, visando a aplicação em uma situação real que está sendo estudada ao redor de todo o mundo. Diante de diversas comparações entre o uso de transportes públicos e o aumento dos casos de Covid-19, foram selecionadas duas bases de dados reais (uma sobre transporte e outra sobre casos de Covid-19) para verificar a correlação entre as variáveis. Nesta aplicação, as duas bases serão tratadas como séries temporais regulares, abstraindo-se qualquer característica epidemiológica dos dados relacionados às contaminações.

Este capítulo apresenta um método de análise e sua aplicação prática para analisar dados temporais de transportes públicos e casos de COVID-19 no estado do Rio de Janeiro. Para tanto, será utilizada a linguagem R com todas as ferramentas necessárias para as análises que serão realizadas. Ao final do capítulo, o leitor terá uma experiência prática com o R que pode ser adaptada e reproduzida em outros cenários. Bem como obterá informações sobre a COVID-19 e sua relação com o transporte público o que, por sua vez, pode proporcionar insights ao leitor para se aprofundar no tema por meio de pesquisas e outras análises.

Durante o desenvolvimento, serão apresentadas ferramentas de visualização e manipulação de dados, juntamente com o procedimento empregado em cada função, bem como os resultados encontrados.

8.2. Metodologia

Nesta seção serão apresentados os principais passos e escolhas para análise. A cada etapa apresentada, também serão abordados os conceitos básicos necessários para sua compreensão. A Figura 8.1 apresenta as fases da pesquisa que será desenvolvida neste capítulo.

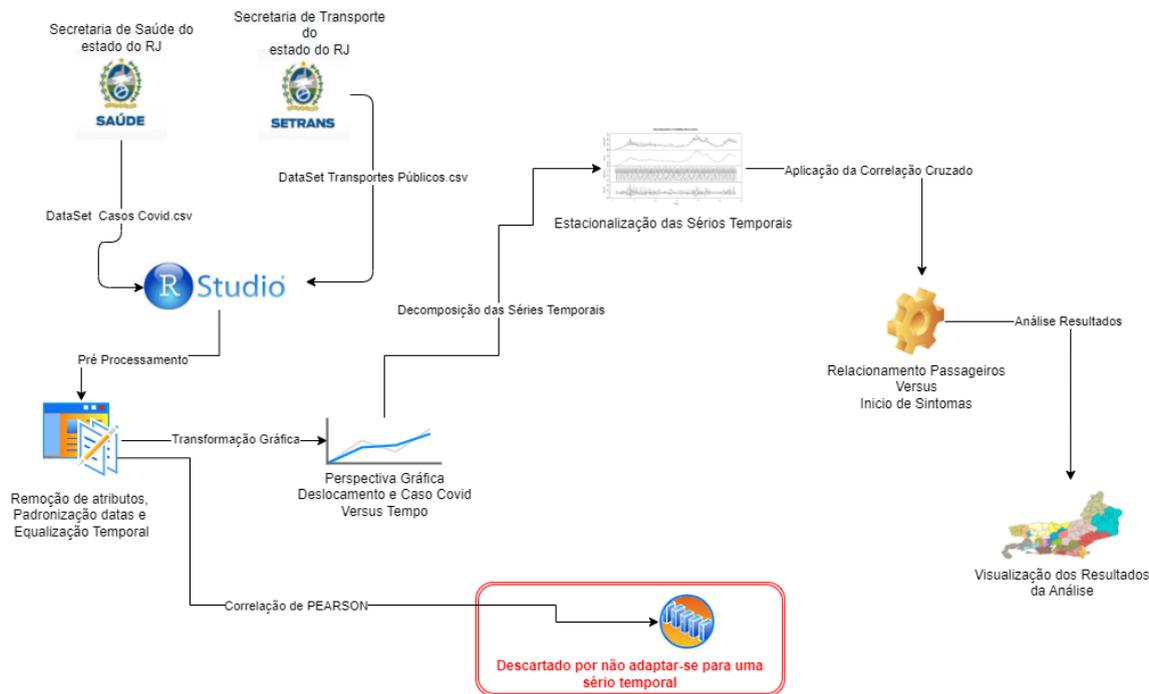


Figura 8.1. Fases da Pesquisa (Fonte: Autores)

8.2.1 Séries temporais

Segundo Ehlers (2007), uma série temporal consiste em um conjunto de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados está no fato de que as observações vizinhas são dependentes, tornando necessário que essas dependências sejam levadas em consideração durante as análises e modelagens. Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante, para a análise em séries temporais a ordem dos dados é crucial. Vale ressaltar também que o tempo pode ser substituído por outras variáveis como espaço, profundidade, etc.

Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes, o estudo de séries temporais requer o uso de técnicas específicas. Dados de séries temporais surgem em vários campos do conhecimento tais como: economia (preços diários de ações, taxa mensal de desemprego, produção industrial); medicina (eletrocardiograma, eletroencefalograma); epidemiologia (número mensal de novos casos da doença); meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento); etc.

Ehlers (2007) definiu algumas características particulares a este tipo de dados, por exemplo:

- Observações correlacionadas são mais difíceis de serem analisadas e, portanto, requerem técnicas específicas;
- É necessário levar em conta a ordem temporal das observações;
- Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover;

- A seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação; e
- É mais difícil lidar com observações perdidas e dados discrepantes devido à natureza sequencial.

De um modo geral, os principais objetivos que podem ser alcançados ao se estudar séries temporais estão dispostos a seguir:

- Descrição: onde pretende-se descrever propriedades da série, tais como o padrão de tendência, a existência de variação sazonal ou cíclica, observações discrepantes (*outliers*), alterações estruturais como mudanças de padrão, etc.
- Explicação: onde se utiliza a variação em uma das séries para explicar o comportamento da outra.
- Predição: onde se objetiva prever valores futuros com base em valores passados. Neste caso, assume-se que o futuro envolve incerteza, o que impede previsões exatas, porém se desenvolvem os estudos visando a máxima minimização dos erros de previsão.
- Controle: nos casos em que os valores da série temporal medem a “qualidade” de um processo e o objetivo está em controlar o mesmo. Como exemplo, pode-se citar o controle estatístico de qualidade onde as observações são representadas em cartas de controle.

Ainda de acordo com Ehlers (2007), uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Todavia, a maior parte das séries na prática apresentam alguma forma de não estacionariedade. Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, faz-se necessária a transformação dos dados originais antes da realização das análises, para tal, deve ser removida a tendência e a sazonalidade.

A tendência está autoexplicativamente relacionada com a tendência da série crescer ou decrescer ao longo do tempo. Já a sazonalidade está associada a um comportamento que tende a se repetir a cada período x de tempo em uma série temporal. A sazonalidade se apresenta em dois tipos, são eles:

- Aditiva: a série apresenta flutuações sazonais mais ou menos constantes não importando o nível global da série.
- Multiplicativa: o tamanho das flutuações sazonais varia dependendo do nível global da série.

8.2.2 Correlação

Neste estudo, pretende-se correlacionar os dados de utilização do transporte público com os casos de Covid-19 no Estado do Rio de Janeiro. Em geral, é comum para a análise de correlações se utilizar a correlação de Pearson. De acordo com Figueiredo Filho e Silva Junior (2009) essa forma de correlação se caracteriza como uma medida de associação linear entre variáveis. Em termos estatísticos, duas variáveis se associam

quando elas guardam semelhanças na distribuição dos seus escores, podendo se associar a partir da distribuição das frequências ou pelo compartilhamento de variância. No caso da correlação de Pearson, ela é uma medida da variância compartilhada entre duas variáveis. Por outro lado, o modelo linear supõe que o aumento ou decréscimo de uma unidade na variável X gera o mesmo impacto em Y. Em termos gráficos, por relação linear entende-se que a melhor forma de ilustrar o padrão de relacionamento entre duas variáveis é através de uma linha reta. Portanto, a correlação de Pearson exige um compartilhamento de variância e que essa variação seja distribuída linearmente.

O formato dos dados com observações diárias caracteriza uma série temporal, onde não cabe a aplicação da Correlação de Pearson, uma vez que este modelo considera a correlação linear entre duas variáveis, dificultando uma análise em dias diferentes, como é o caso das bases de dados a serem estudadas. Neste caso, o procedimento adequado se dá com a aplicação da correlação cruzada que permite identificar correlações que acontecem em qualquer período de tempo.

De acordo com Silva Filho (2014), o método de correlação cruzada é um método estatístico capaz de estimar o expoente que caracteriza a correlação de longo alcance entre duas séries temporais, em regime não estacionário. Dessa forma, a função de correlação cruzada estima a correlação entre duas séries temporais tendo o tempo incluído como uma variável. Dessa forma, é possível estimar a influência de uma série na outra mesmo que esta não tenha sido gerada no mesmo tempo, como por exemplo uma série alocada no tempo t influenciando no comportamento de uma série alocada no tempo $t + k$. O atraso ou defasagem é chamado de lag e no exemplo anterior está denotado como k , podendo adquirir valores positivos ou negativos. Cabe ressaltar que para a aplicação da função de correlação cruzada a série deve ser estacionária, sendo necessária portanto a remoção de quaisquer tendência e sazonalidade presentes em ambas as séries.

8.3 Fonte de Dados

A Secretaria de Transportes do estado do Rio de Janeiro (Setrans) é o órgão responsável por realizar estudos, pesquisas e planejamento do sistema de transportes do estado, bem como operar adequadamente os serviços de transportes e de terminais rodoviário de passageiros, metroviário, ferroviário e hidroviário. A Setrans disponibilizou para esta pesquisa dados de quantidade de passageiros diários para todos os modos de transporte por ela operados no período de 09/03/2020 até 11/04/2021.

Os modos de transporte público selecionados para esta análise foram: metrô, trem, barcas, ônibus municipais e ônibus intermunicipais. Além disso, para todos os modos foram considerados os valores totais de passageiros somando todas as linhas/ramais de cada modo.

Visando a comparação dos dados de utilização de transporte público com a evolução da pandemia da Covid-19 para o Estado do Rio de Janeiro, foram coletados dados referentes ao número de casos de Covid-19 no estado para o mesmo período dos dados fornecidos pela Setrans. A fonte dos dados foi a plataforma da Secretaria de Saúde do RJ que fornece uma base em formato csv com dados atualizados diariamente pelo Centro de Informações Estratégicas e Resposta de Vigilância em Saúde (CIEVS-RJ) da Secretaria de Saúde do Estado do Rio de Janeiro, a partir do sistema

esus-VE e SIVEP-Gripe, em articulação com as vigilâncias das secretarias municipais de saúde do Estado.

Buscando maior assertividade dos valores referentes aos números de casos, na tentativa de minimizar os erros referentes a atrasos da data de divulgação, foram selecionados os casos confirmados por data de início dos sintomas. A plataforma permite selecionar os parâmetros para gerar diferentes tabelas, as quais podem ser baixadas em formato csv. Para esta análise foi selecionado “Data de início dos sintomas” como parâmetro de linhas e “Casos confirmados” para as medidas.

Ao final da etapa de coleta de dados, foram geradas duas planilhas em formato csv, são elas: “12.04 - Dados.csv” para os dados de transporte público e "inicio_x_casos.csv" para dados de Covid-19.

O código completo e a base de dados de referência para os dados de casos de Covid-19 estão disponíveis no Github

<<https://github.com/juliafagundesoc/Cap-tulo-8-SBC>>

8.3.1 Importação dos dados

A IDE (*integrated development environment*) utilizada foi o RStudio que pode ser utilizada tanto para R quanto para Python. Em um primeiro momento, ambos os arquivos csv devem ser salvos na mesma pasta, a qual será definida como o diretório no RStudio como indicado na Figura 8.2.

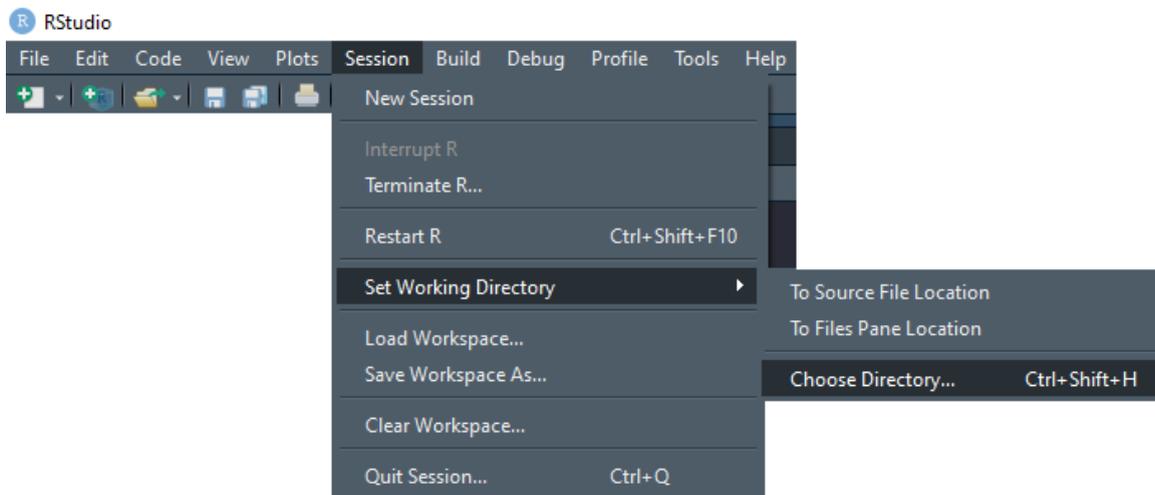


Figura 8.2. R Studio Seleção de Diretório (Fonte: Autores)

A Figura 8.3 ilustra a importação das bases de dados incluindo os parâmetros para adequação dos *Datasets*. Nesta etapa, foi selecionado o símbolo “;” como separador das colunas para a conversão do csv, além disso, o padrão de codificação (*encoding*) foi definido como “Latin1” para evitar erros com a presença de acentos nos nomes das variáveis. Nesta etapa também cabe incluir a definição do símbolo de separação de decimais, no caso da base enviada pela Setrans era a vírgula, por esse motivo o parâmetro “decimal_mark” foi definido como “;”. Ao final, os tipos das colunas que estavam pré-definidas incorretamente (estavam como caracter

e eram numéricas) foram corrigidos utilizando o parâmetro “col_types”, igualando as respectivas colunas a “col_number()” para defini-las como tipo numéricas.

```

inicio_x_casos <- read_delim("inicio_x_casos.csv",
                             ";", escape_double = FALSE,
                             locale = locale(encoding = "Latin1"),
                             trim_ws = TRUE)

X12_04_Dados <- read_delim("12.04 - Dados.csv",
                             ";", escape_double = FALSE,
                             locale = locale(decimal_mark = ",", encoding = "Latin1"),
                             trim_ws = TRUE,
                             col_types = cols(Metro_Tot = col_number(),
                                                Trem_Tot = col_number(),
                                                Barcas_Tot = col_number(),
                                                RMRJ = col_number(),
                                                Intermunicipais = col_number()))

```

Figura 8.3. Trecho do código importação das bases csv de Casos e Transportes Públicos (Fonte: Autores)

8.4. Desenvolvimento

Em um primeiro momento, cabe ressaltar que o R disponibiliza a documentação de cada função utilizando “?” antes da mesma, nesta é possível ter acesso às informações de modo de usar, parâmetros, aplicações, além de apresentar exemplos de aplicação.

No desenvolvimento deste estudo, de acordo com o contexto da proposta de ensinar a linguagem R, em uma aplicação prática, foram selecionadas as seguintes bibliotecas:

- **Dplyr** é o pacote mais útil para realizar transformação de dados, aliando simplicidade e eficiência de uma forma elegante. Os scripts em R que fazem uso inteligente dos verbos ‘dplyr’ e as facilidades do operador pipe tendem a ficar mais legíveis e organizados sem perder velocidade de execução. São suas principais funções: *select()* - seleciona colunas; *arrange()* - ordena a base; *filter()* - filtra linhas; *mutate()* - cria/modifica colunas; *group_by()* - agrupa a base; e *summarise()* - sumariza a base [Battisti e Smolski, 2019].
- **Lubridate** é o pacote que traz diversas funções para extrair os componentes de um objeto da classe date. Principais funções: *second()* - extrai os segundos; *minute()* - extrai os minutos; *hour()* - extrai a hora; *wday()* - extrai o dia da

semana.; `mday()` - extrai o dia do mês; `month()` - extrai o mês; e `year()` - extrai o ano [Battisti e Smolski, 2019].

- **Ggplot2** é o pacote que constrói um gráfico camada por camada. Este pacote permite criar o nosso canvas, um quadro em branco onde vamos colocar todas as outras camadas do gráfico. Dentro do R, isso corresponde a uma lista com as informações necessárias para a criação do gráfico. Cada camada adicionada ao `ggplot` adiciona mais informações nessa lista [Battisti e Smolski, 2019].
- **Ggcorrplot** é o pacote utilizado para visualizar facilmente uma matriz de correlação usando 'ggplot2'. Ele fornece uma solução para reordenar a matriz de correlação e exibe o nível de significância no gráfico. Também inclui uma função para calcular uma matriz de valores “p” de correlação [Battisti e Smolski, 2019].
- **Stats** é um pacote que facilita a visualização da correlação entre duas ou mais camadas [Battisti e Smolski, 2019].
- **Zoo** é um pacote para observações indexadas totalmente ordenadas. Destina-se particularmente a séries temporais irregulares de vetores / matrizes e fatores numéricos. Os principais objetivos do projeto do zoo são a independência de uma classe de índice / data / hora específica e consistência com ts e R de base, fornecendo métodos para estender os genéricos padrões [Battisti e Smolski, 2019].

8.4.1. Pré-processamento

Na etapa de pré-processamento, apresentada na Figura 8.4, foram removidas algumas observações e colunas obsoletas, bem como a adequação dos nomes das colunas para que os dois *Dataframes* ficassem harmonizados e com uma perspectiva comum. Além disso, foi incluída a variável “Municipais” que engloba a parcela dos ônibus que não fazem trajeto intermunicipal. Esta inclusão fez-se necessária para que não exista duplicidade de dados referentes aos ônibus, uma vez que a variável “RJMJ” englobava a soma dos ônibus municipais e intermunicipais. A variável “Total” também foi incluída, contemplando a soma de todos os modos de transportes.

Como a base de dados de casos de Covid-19 apresentava algumas datas ausentes e a característica dos dados não justificava imputação desses valores, optou-se por remover estas datas do *Dataframe* final denominado “df”. Ainda de acordo com a característica dos dados não foram removidos potenciais outliers.

```

# Limpeza e ajuste do Dataframe de casos de Covid-19
casos=inicio_x_casos[-c(1,597),]
colnames(casos)[1] <- "Data"
colnames(casos)[2] <- "Casos"
casos$Data <- as.Date(casos$Data)
# Limpeza e ajuste do Dataframe de transportes
df=X12_04_Dados[-c(400:615),-c(1,4:6,8:17,19:26,28:31,34:57)]
df <- rename(df, Data = Dia)
df$Data <- as.Date(df$Data)
df["Municipais"] <- df$RMRJ - df$Intermunicipais
df["Total"] <- df$Metro_Tot + df$Trem_Tot + df$Barcas_Tot + df$RMRJ
df <- df[,-6]
# Remoção das datas com valores ausentes de número de casos
data = df$Data
data <- as_data_frame(data)
colnames (data)[1] <- "Data"
x = left_join(data,casos,by= "Data")
df["Casos"] <- x$Casos

```

Figura 8.4 Trecho do código limpeza e ajustes da base (Fonte: Autores)

A Figura 8.5 apresenta as 10 primeiras observações de “df” que apresenta 399 observações no total. Além das variáveis dos modos de transporte e de casos, a variável “Tipo_de_dia” assume valores de D.U. referente aos dias úteis, além de “Fim_de_semana”, “Feriado” e “Enforcado”.

	Data	Tipo_de_dia	Metro_Tot	Trem_Tot	Barcas_Tot	Intermunicipais	Municipais	Casos
1	2020-03-09	D.U.	887593	605061	77532	965632	2907782	42
2	2020-03-10	D.U.	901063	614043	77315	983721	2947688	154
3	2020-03-11	D.U.	918853	605926	73641	948307	2861275	66
4	2020-03-12	D.U.	879675	602627	75166	964846	2909712	74
5	2020-03-13	D.U.	903167	604208	75639	962760	2925060	112
6	2020-03-14	Fim_de_semana	407462	235571	24799	591606	1860578	127
7	2020-03-15	Fim_de_semana	228293	96007	11651	322015	1104619	306
8	2020-03-16	D.U.	581026	465951	51343	839176	2316611	220
9	2020-03-17	D.U.	448427	395721	34190	769230	2015355	201
10	2020-03-18	D.U.	360805	339936	25160	690204	1767611	214

Figura 8.5. Primeiras observações da base gerada (Fonte: Autores)**8.4.2. Visualizações**

A função `ggplot()` da biblioteca “`ggplot2`” permite elaborar um gráfico de linhas com todas as variáveis utilizando o parâmetro “`geom_line`”, utilizado para a elaboração de gráficos de linhas, a aplicação da função está apresentada na Figura 8.6.

```
#Evolução Diária por modo de transporte

ggplot(df, aes(x=Data, y=Metro_Tot))+
  geom_line(aes(col= "Metrô"))+
  geom_line(aes(y=Trem_Tot, col= "Trem"))+
  geom_line(aes(y=Barcas_Tot, col= "Barcas"))+
  geom_line(aes(y=Municipais, col= "Ônibus Municipais"))+
  geom_line(aes(y=Intermunicipais, col= "Ônibus Intermunicipais"))+
  geom_line(aes(y=Casos, col= "Casos"))+
  theme_classic()+
  labs(x="Data",
       y="Modo de Transporte",
       color=NULL)+
  theme(legend.position = "top")
```

Figura 8.6. Trecho do código Evolução diária por modo de transporte

A Figura 8.7 apresenta a plotagem do gráfico gerado com a aplicação da função anterior, neste é possível observar que a diferença de escala das variáveis dificulta a comparação entre o comportamento das mesmas. As variáveis “Casos” e “Barcas” por exemplo apresentam uma variação quase imperceptível em comparação com as demais variáveis, o que não condiz com a realidade.

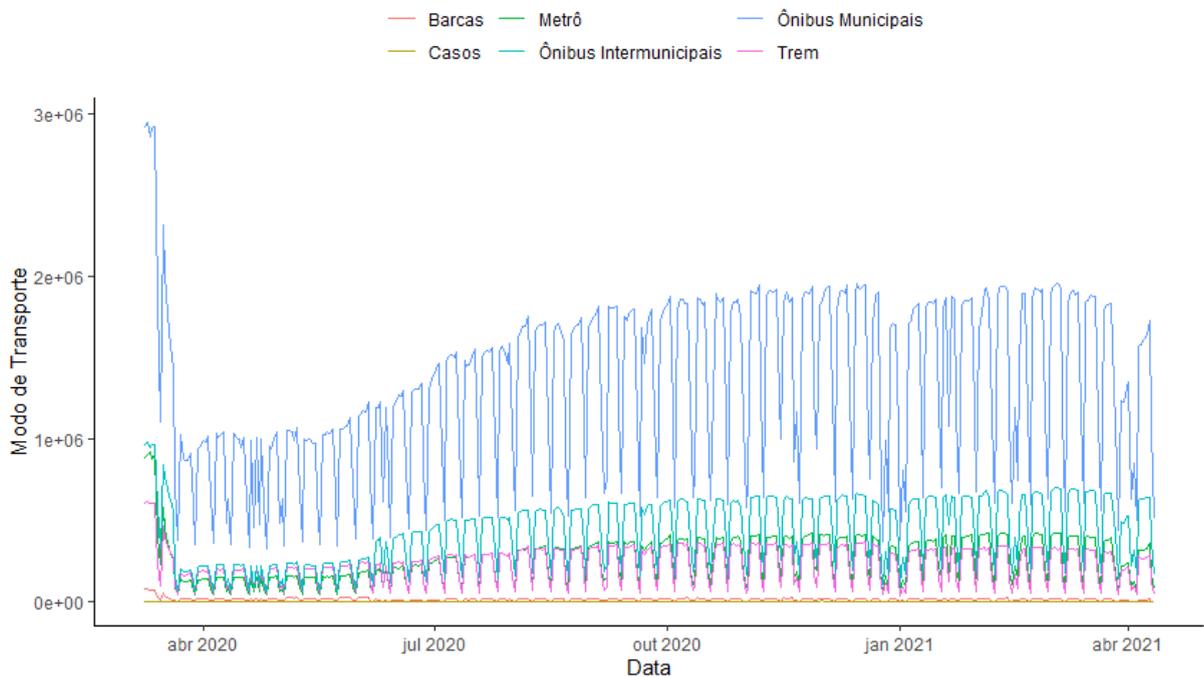


Figura 8.7. Evolução diária das variáveis

Para padronização dos gráficos de forma que seja possível analisar o comportamento dos mesmos foi aplicada a função `scale()` da base (Figura 8.8) para todas as variáveis exceto “Data”, esta função executa o procedimento de centralizar em zero e alterar a escala para desvio padrão.

```
df_scale <- df
df_scale = data_frame("Data"      = df_scale$Data,
                     "Metro_Tot"  = scale(df_scale$Metro_Tot),
                     "Trem_Tot"   = scale(df_scale$Trem_Tot),
                     "Barcas_Tot" = scale(df_scale$Barcas_Tot),
                     "Municipais" = scale(df_scale$Municipais),
                     "Intermunicipais" = scale(df_scale$Intermunicipais),
                     "Casos"      = scale(df_scale$Casos)
                    )
```

Figura 8.8. Trecho do código Padronização das variáveis

A função `ggplot()` permite ainda a inclusão de marcos no gráfico. A Figura 8.9 apresenta a aplicação da função para os dados padronizados, incluindo ainda um marco representado por uma linha vertical, aplicado com a utilização do parâmetro “`geom_vline`” seguido do parâmetro “`geom_text`” que inclui a legenda no marco.

```

# Evolução diária padronizada para estudo do comportamento do gráfico
ggplot(df_scale, aes(x=Data, y=Metro_Tot))+
  geom_line(aes(col= "Metrô"))+
  geom_line(aes(y=Trem_Tot, col= "Trem"))+
  geom_line(aes(y=Barcas_Tot, col= "Barcas"))+
  geom_line(aes(y=Municipais, col= "Ônibus Municipais"))+
  geom_line(aes(y=Intermunicipais, col= "Ônibus Intermunicipais"))+
  geom_line(aes(y=Casos, col= "Casos"))+
  geom_vline(data = subset(df_scale, Data == "2020-03-16"),
            aes(xintercept = Data),    size = 0.3, colour = "black")+
  geom_text(data=subset(df_scale, Data == "2020-03-16"),
            mapping=aes(x=Data, y=0,  label= "Decretada situação de emergência"),
            size=4, angle=90, vjst=1, hjust=-0.7) +
  theme_classic()+ labs(x="Data",
                        y="Modo de Transporte",
                        color=NULL)+
  theme(legend.position = "top")

```

Figura 8.9 - Evolução diária padronizada para estudo do comportamento gráfico

O resultado da plotagem com os dados padronizados está apresentado na Figura 8.10. Neste, é possível observar uma queda bem definida após o início da pandemia, além disso, identifica-se um padrão semanal bem definido para todos os modos de transportes, com quebras pontuais que coincidem com feriados e recessos. Já os dados referentes aos casos de Covid-19 apresentam um comportamento mais caótico, onde visualmente não é possível identificar um padrão.

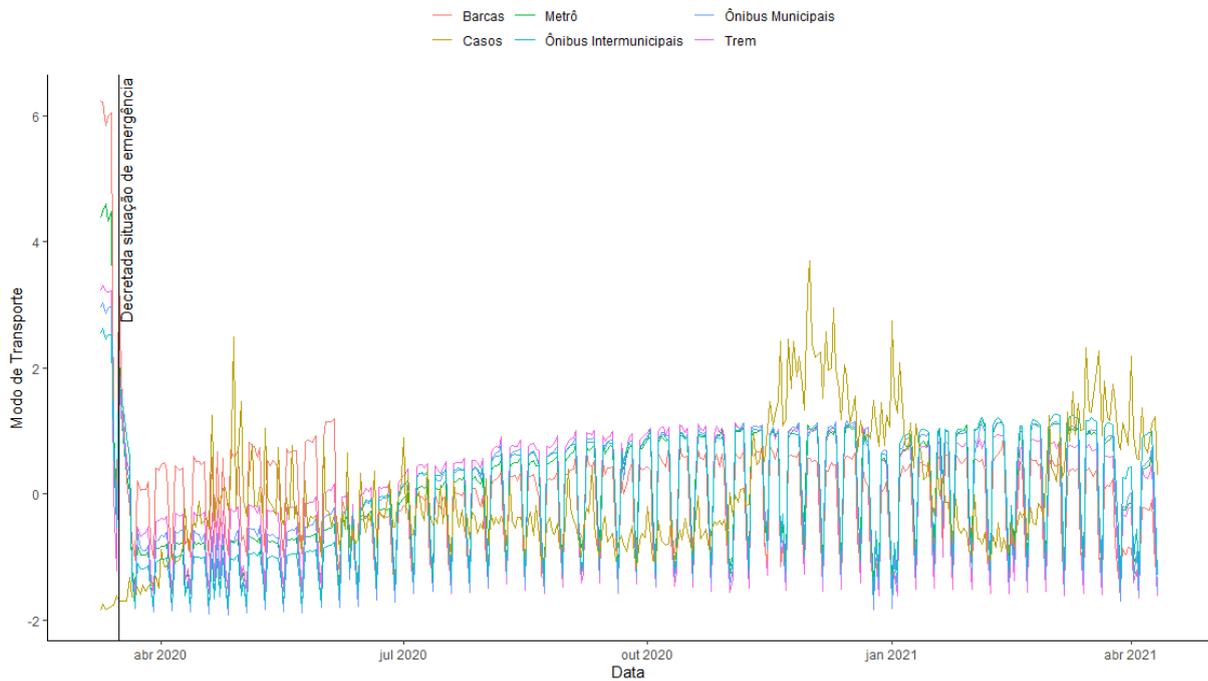


Figura 8.10. Gráfico de evolução diária com variáveis padronizadas

Em uma análise utilizando dados que não se apresentam no formato de uma série temporal, seria possível elaborar uma matriz de correlação entre todas as variáveis utilizando a Correlação de Pearson. Apesar de não ser adequada para este exemplo, esta será aplicada apenas a título de ilustração. A Figura 8.11 apresenta o procedimento para a criação da matriz, para tal, utiliza-se a função `cor()` da biblioteca “stats” para cálculo das correlações e, em seguida, estas são aplicadas na função `ggcorrplot()` da biblioteca “ggcorrplot” para a elaboração da matriz.

```

# Elaboração da Matriz de Correlação de Pearson

x = df_scale[,c("Metro_Tot", "Trem_Tot", "Barcas_Tot", "Municipais",
               "Intermunicipais", "Casos")]

matriz <- round(cor(x), 2)
ggcorrplot(matriz, type = "lower",
            lab = TRUE,
            lab_size = 3,
            colors = c("firebrick", "white", "dodgerblue4"),
            title = "Matriz de Correlação de Pearson",
            ggtheme = theme_classic)

```

Figura 8.11. Trecho do código Criação da Matriz de Correlação Pearson

A Figura 8.12 apresenta o resultado da matriz de correlação, nesta é possível observar uma correlação forte entre todos os modos de transportes, como era de se esperar pelo comportamento visual do gráfico disposto na Figura 8.10. É possível supor que a correlação de Pearson ilustra bem a correlação dos modos de transportes entre si porque espera-se que esta ocorra no mesmo dia, sem defasagem de tempo. Por outro lado, a correlação encontrada entre o número de casos e todos os modos de transportes apresenta-se abaixo de 0,2, o que em geral se considera um resultado muito fraco para ser considerado. Neste caso, não é possível esperar um resultado fidedigno da correlação de Pearson uma vez que uma variável pode influenciar na outra em tempos diferentes, por exemplo, uma pessoa infectada em um modo de transporte no dia x só deve apresentar sintomas no dia $x + t$, sendo t o tempo desde a infecção até o aparecimento dos sintomas. Dessa forma, cabe a aplicação da correlação cruzada para identificar correlação em qualquer período de tempo, como será apresentado no item 1.3.3.

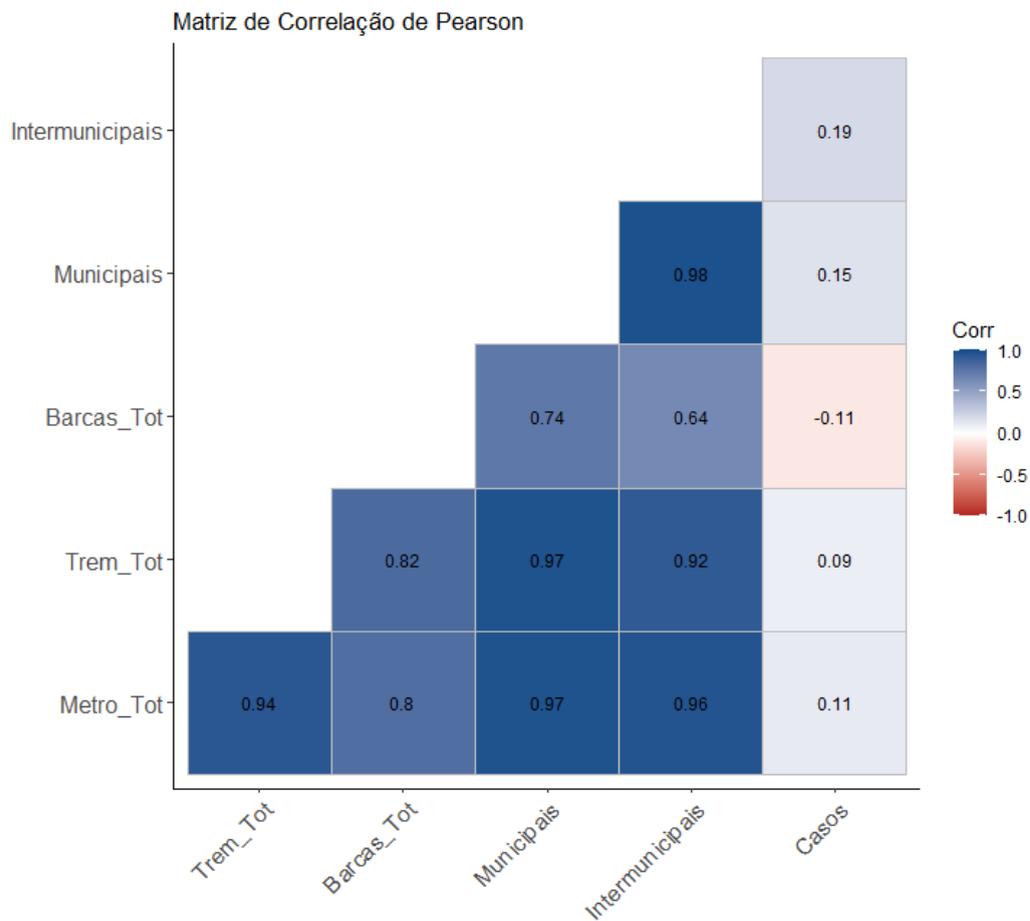


Figura 8.12 - Matriz de Correlação Pearson

8.4.3 Aplicação da Correlação Cruzada

Para a aplicação da correlação cruzada, a variável a ser comparada com o número de casos será “Total”, que engloba a soma por dia de todos os modos de transportes. Por se tratar de uma série temporal, algumas etapas devem ser cumpridas para estacionarização da série de forma que se possa correlacionar as duas variáveis de maneira adequada. Em um primeiro momento, utiliza-se a função `ts()` para criar o objeto de série temporal como disposto na Figura 8.13, as variáveis utilizadas para comparação serão “Total” que engloba a soma de todos os modos de transportes e “Casos”. Como a base de dados está organizada por dia, a frequência da série temporal, indicada pelo parâmetro “frequency”, deve ser definida por 7, indicando um período natural de uma semana. O resultado da plotagem das duas séries temporais está apresentado na Figura 8.14.

```
# Séries Temporais

series <- df[,c("Data", "Total", "Casos")]

ts_total = ts(series$Total,frequency = 7)
ts_casos = ts(series$Casos,frequency = 7)

plot(cbind(ts_total,ts_casos), main="Séries Temporais")
```

Figura 8.13. Trecho do Código Séries Temporais

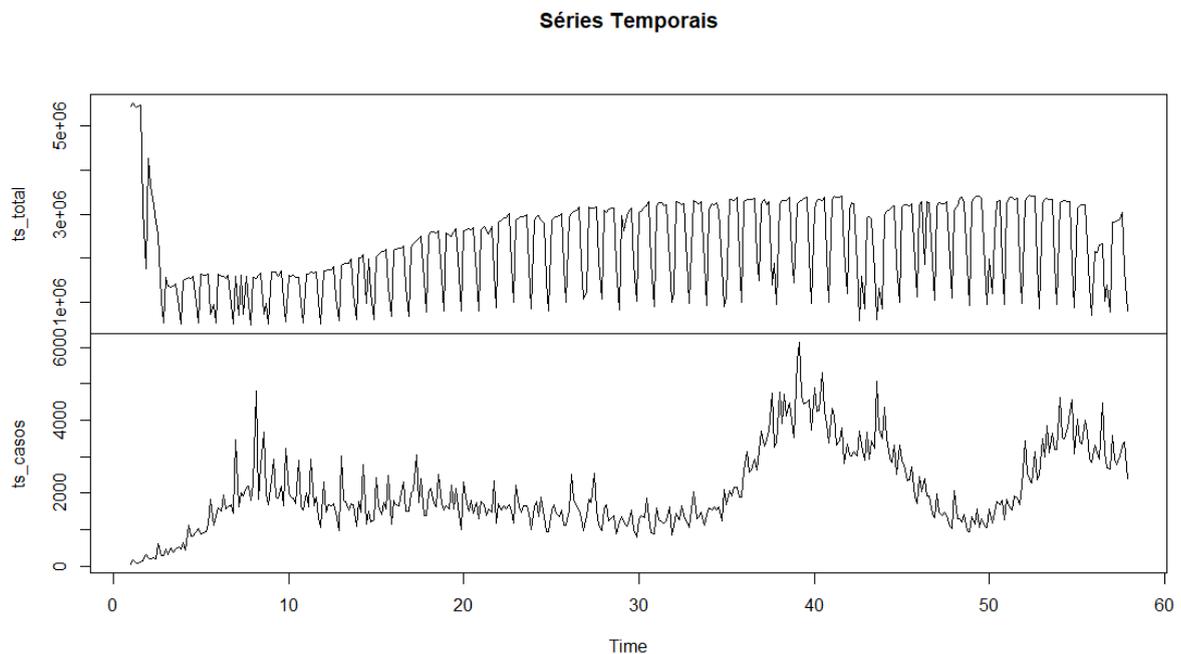


Figura 8.14. Gráfico Séries Temporais

A etapa seguinte consiste decompor as séries temporais com o objetivo de separar a tendência e a sazonalidade das mesmas, para tal, pode ser aplicada a função `decompose()` da biblioteca “stats” como apresentado na Figura 8.15. Como a variação sazonal apresenta constância ao longo do tempo, o parâmetro “type” recebe o argumento 'additive' relativo ao modelo aditivo, do contrário este deve receber o argumento 'multiplicative' referente ao modelo multiplicativo onde a variação sazonal aumenta ao longo do tempo.

```

# Remoção de tendência e de sazonalidade

decompose_total = decompose(ts_total, type = 'additive')
plot(decompose_total)
r_total = decompose_total$random

decompose_casos = decompose(ts_casos, type = 'additive')
plot(decompose_casos)
r_casos = decompose_casos$random

plot(cbind(r_total,r_casos), main=" ")

```

Figura 8.15. Trecho do Código remoção de tendência e sazonalidade

As Figuras 8.16 e 8.17 apresentam ,respectivamente, os resultados das plotagens da função `decompose()` aplicada à série temporal relativa ao total de transportes públicos e à série temporal relativa aos casos de Covid-19. A saída da função apresenta 4 gráficos, o primeiro denominado “observed” apresenta a série temporal original, o segundo denominado “trend” apresenta a tendência da série, o gráfico “seasonal” representa a sazonalidade da mesma e o último, “random” consiste no resultado final já randomizado com a remoção da tendência e da sazonalidade. Dessa forma, a série gerada como “random” para cada variável (Figura 8.18) passará a ser a série de entrada para a correlação, as séries randomizadas para “Total” e “Casos” foram armazenadas nas variáveis “r_total” e ‘r_casos”, respectivamente.

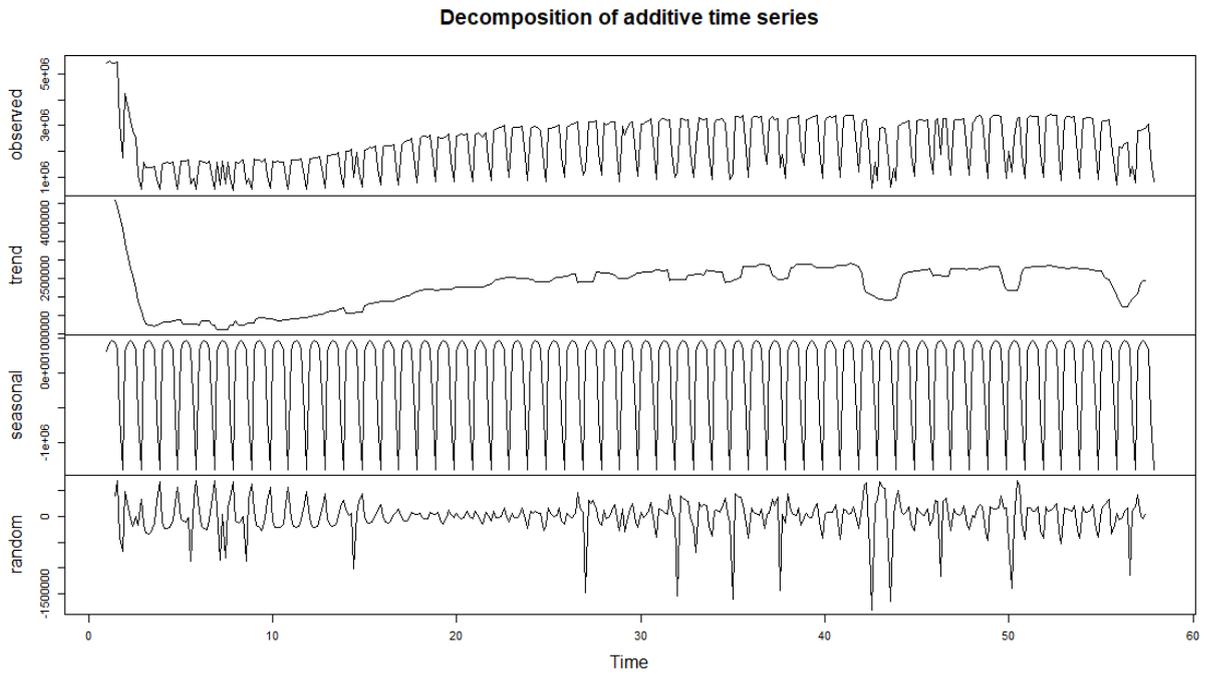


Figura 8.16. Gráfico Decomposição variável Total

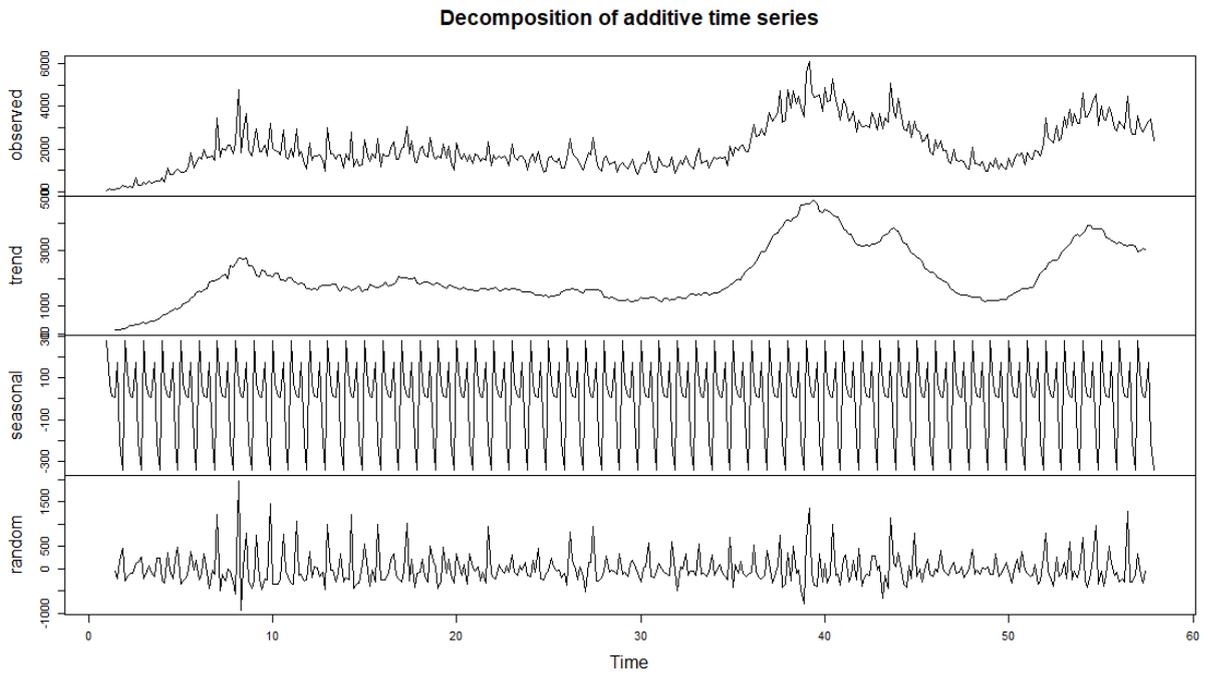


Figura 8.17. Gráfico Decomposição variável Casos

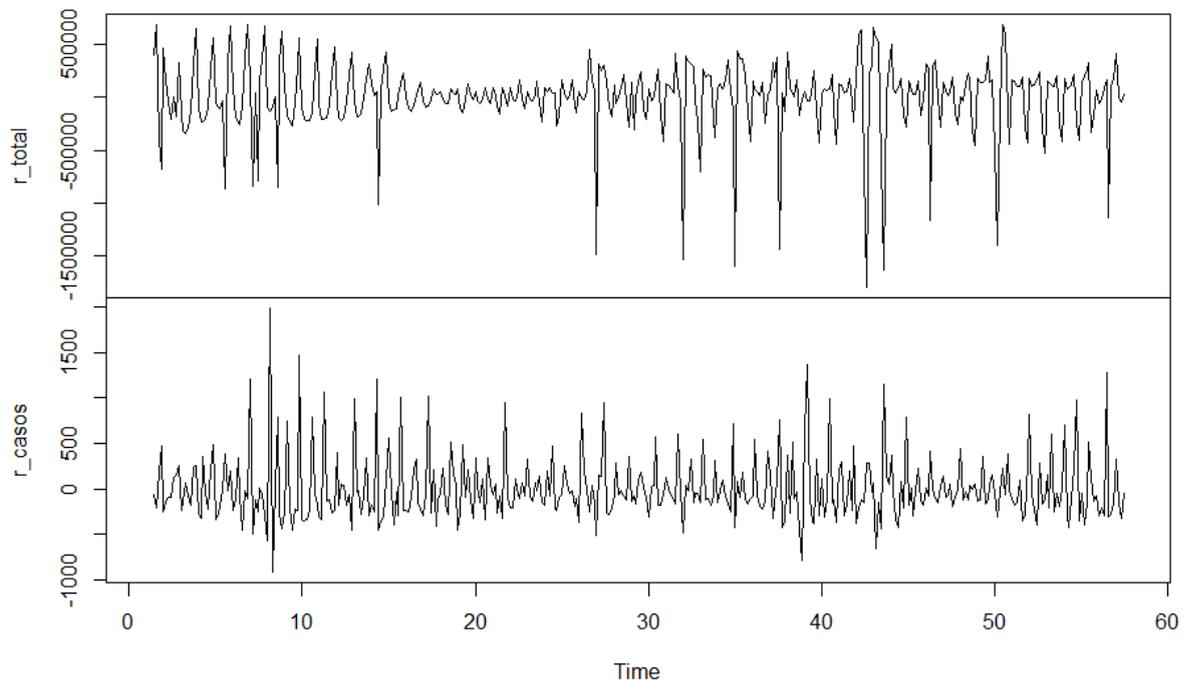


Figura 8.18. Gráfico Séries temporais Randomizadas

Como as variáveis a serem estudadas são séries temporais, a correlação mais adequada a ser aplicada é a correlação cruzada, nesta é possível identificar a correlação para diferentes *lags* (atrasos). No R, a função “*ccf ()*” da biblioteca “*stats*” permite a análise da correlação cruzada, nesta, as entradas assumem o formato *ccf (x,y)*, onde *x* representa a variável líder e *y* a variável dependente. Nesta aplicação, em um primeiro momento, a variável *r_total* referente aos transportes públicos foi definida como líder e a variável *r_casos*, referente aos casos de Covid-19, como dependente (esta hipótese poderá ser confirmada após o resultado da correlação como será apresentado a seguir). A Figura 8.19 apresenta a aplicação da função *ccf ()* seguida da definição do *lag* e correlação máximos, bem como da conversão do *lag* para o número de dias.

```

# Correlação Cruzada

ccf = ccf(r_total,r_casos, type = "correlation",
          na.action = na.pass,
          ylab = "Correlação",
          main = "Correlação Cruzada")

cor = ccf$acf[,,1]
lag = ccf$lag[,,1]
res = data.frame(cor,lag)
lag_max = res[which.max(res$cor),]$lag
lag_max
cor_max = res[which.max(res$cor),]$cor
cor_max
dias = lag_max *7
dias

```

Figura 8.19. Trecho do Código Correlação Cruzada

A Figura 8.20 apresenta o resultado da correlação cruzada, as linhas tracejadas azuis representam o intervalo de confiança calculado automaticamente pela função `ccf()` em função do tamanho da amostra e o *lag* máximo, os valores de correlação para além desse intervalo são considerados significativos. O *lag* negativo indica que a definição de variável líder foi correta, de forma que os valores referentes ao início de sintomas de covid estão relacionados com o uso do transporte público em um tempo anterior. Se o *lag* máximo fosse positivo, indicaria que a hipótese estava incorreta e as variáveis definidas inicialmente como líder e dependente deveriam ser invertidas. A correlação máxima foi de aproximadamente 0,12 identificada em um *lag* = - 0.8571429, como os intervalos são de 7 dias, o valor do *lag* em dias corresponde a - 6 dias.

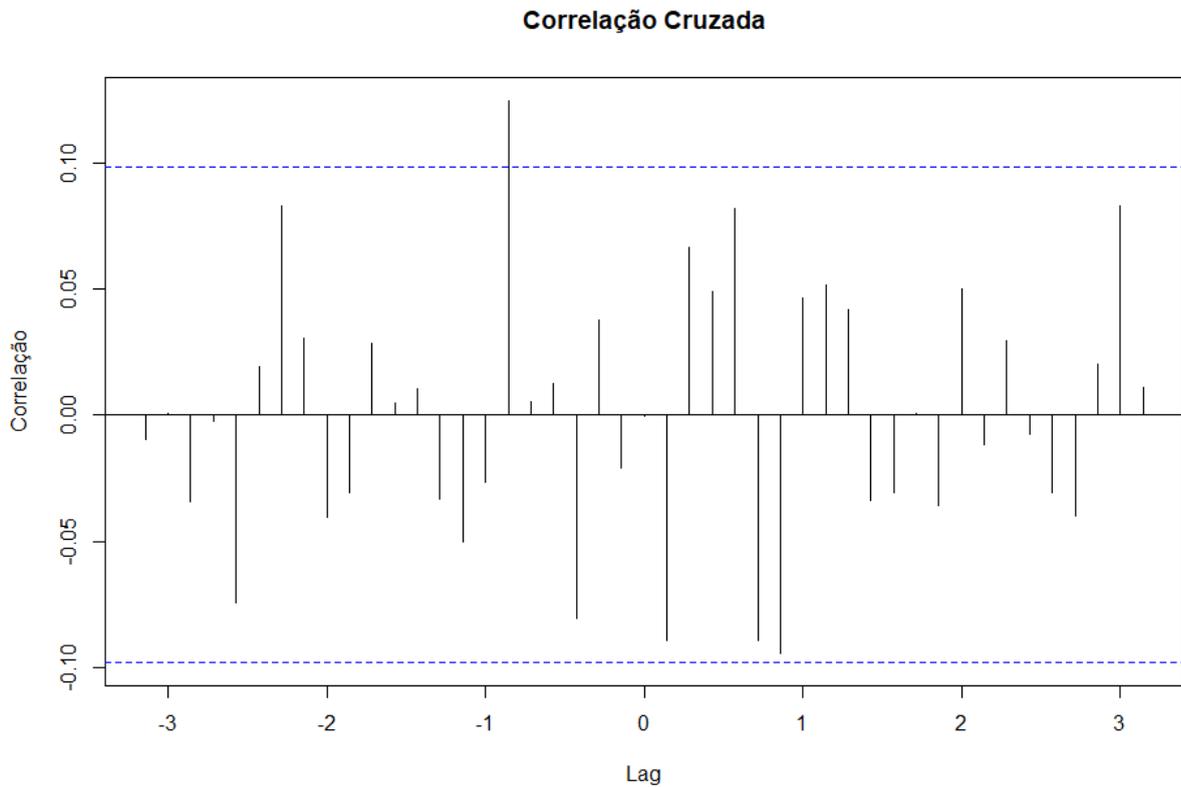


Figura 8.20. Gráfico Correlação Cruzada pós Estacionarização

A título de comparação, a Figura 8.21 apresenta a correlação aplicada entre as séries temporais antes da remoção da sazonalidade e da tendência, nesta pode-se observar diversos valores de correlação em diferentes *lags* tanto positivos quanto negativos, o que indica que a tendência e a sazonalidade das séries estão poluindo os dados, impedindo uma estimativa adequada das correlações. Cabe ressaltar que neste caso os *lags* máximos são positivos, o que acarretaria em uma interpretação incorreta da variável líder, reiterando a importância de uma série estacionária para aplicação da correlação cruzada.

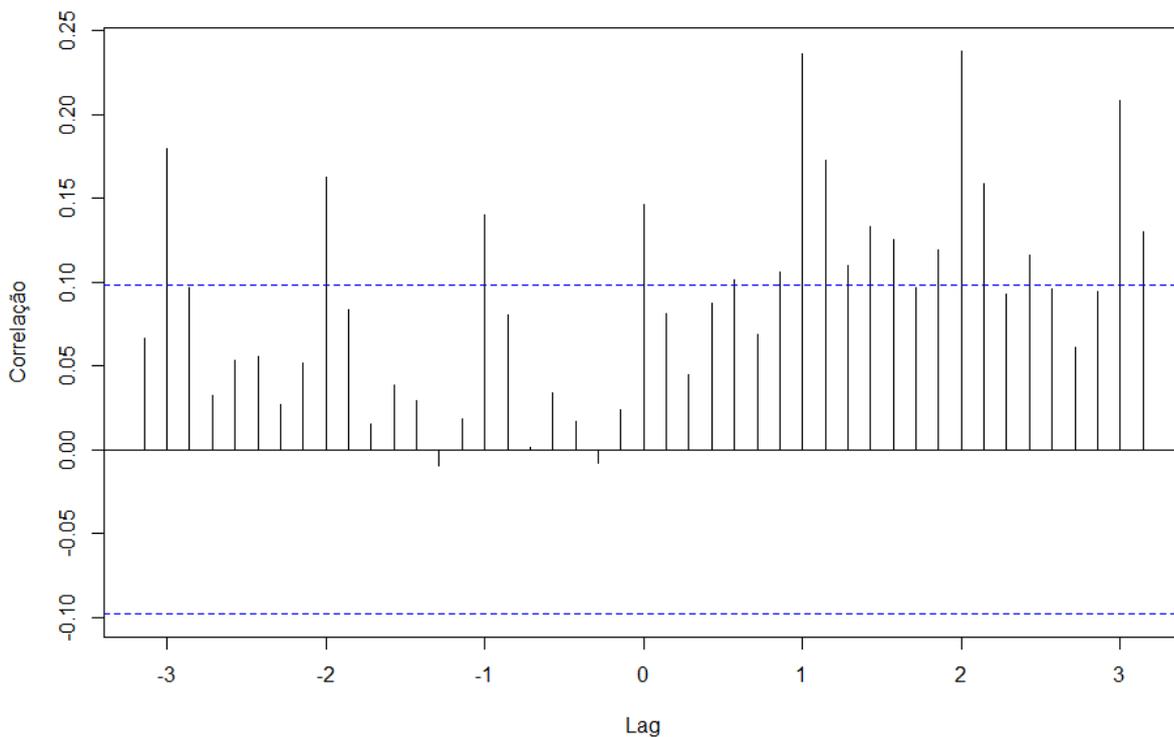


Figura 8.21. Gráfico de Correlação Cruzada com Tendência e Sazonalidade

Os resultados da Figura 8.21 indicam que existe uma correlação entre os dados de utilização de transportes públicos e os dados de contaminações por Covid-19, o *lag* indica que esta correlação de dá com 6 dias de antecedência, ou seja, um aumento de utilização de transportes públicos influencia no aumento do número de casos 6 dias depois. Em uma projeção para a realidade, algumas pesquisas de identificação do padrão de evolução da pandemia estimam que os sintomas da Covid-19 podem aparecer de 2 a 14 dias após o contágio, onde a maioria se encaixa no intervalo de 5 a 7 dias, dessa forma, o atraso de 6 dias encontrado está de acordo com o intervalo esperado.

8.5.Considerações Finais

Este capítulo teve como objeto principal apresentar o uso da linguagem R para análise de dados, com aplicação em séries temporais com dados de transportes públicos e casos de Covid-19. Uma vez que utilizamos base de dados reais, foi possível, além demonstrar técnicas estatísticas e como é a estrutura exigida pela ferramenta 'R', apresentamos resultados importantes para tomada de decisão pública.

Na etapa de pré-processamento, ao harmonizar-se os dados das duas bases recebidas em uma perspectiva integrada, informações sazonais como períodos de final de semana e feriados foram retiradas pois ofereciam distorções nos resultados.

Com a distribuição das informações em formato gráfico e a análise do comportamento da evolução temporal identifica-se um padrão semanal bem definido para todos os modos de transportes, com quebras pontuais que coincidem com feriados e recessos. Entretanto, os dados referentes aos casos de Covid-19 apresentam um

comportamento mais caótico, onde visualmente não é possível identificar um padrão. Desta forma, foi descartado a utilização da Correlação de Pearson pois a tendência e a sazonalidade das séries poluíram os dados, impedindo uma estimativa adequada das correlações.

Contudo, os resultados da correlação cruzada indicaram que existe uma correlação entre os dados de utilização de transportes públicos e os dados de contaminações por Covid-19, o *lag* indica que esta correlação de dá com 6 dias de antecedência, ou seja, um aumento de utilização de transportes públicos influencia no aumento do número de casos 6 dias depois. Esta descoberta vai ao encontro dos dados da Secretaria de Saúde do Estado do Espírito Santo que apontam que o período médio de incubação por coronavírus é de 05 dias [SESA, 2020].

Espera-se que este capítulo sirva de fonte para que mais pesquisadores e alunos que trabalham com análise de dados, ciência de dados ou pesquisa em geral construam uma base de conhecimento que facilite a evolução de futuras pesquisas.

8.6. Referências Bibliográficas

- Battist, I. D. E. and Smolski, F. M. S. (2019) “Software R: Análise estatística de dados utilizando um programa livre”, Editora Faith, Bag é, RS, 2019. Disponível em: <<http://www.editorafaith.com.br/ebooks/grat/978-85-68221-44-0.pdf>>.
- Britto Dalson; Filho, Figueiredo; Alexander, José; et al. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r) . Revista Política Hoje, v. 18, n. 1, 2009. disponível em: <<https://periodicos.ufpe.br/revistas/politica hoje/article/viewFile/3852/3156>>.
- EHLERS, Ricardo S. Análise de séries temporais. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná, v. 1, p. 1-118, 2007. Disponível em: <<https://www.icmc.usp.br/pessoas/ehlers/stemp/stemp.pdf>>
- FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). Revista Política Hoje. Disponível em: <<https://periodicos.ufpe.br/revistas/politica hoje/article/viewFile/3852/3156>>.
- Lima, G. C. L. S., Schechtman, R., Brizon, L. C., Figueiredo, Z. M. (2020). Transporte público e COVID-19. O que pode ser feito?. Rio de Janeiro. Centro de Estudos em Regulação e Infraestrutura da Fundação Getúlio Vargas (FGV CERi).
- Organização Mundial da Saúde- OMS (2020). Coronavirus disease 2019 (COVID19)-SITUATION REPORT 51. Acesso em: 08 de março de 2021. Disponível em: <<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf>>
- Phelan, A. L., Katz, R., Gostin, L. O. The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance, JAMA 2020, DOI:10.1001/jama.2020.1097
- Secretaria de Saúde Estado do Espírito Santos - SESA (2020). Boletim Epidemiológico. Acesso em: 18 de julho de 2021. Disponível em: <<https://coronavirus.es.gov.br/>>

Secretaria de Saúde Estado do Rio de Janeiro - SESA-RJ (2021). Boletim Epidemiológico. Acesso em: 30 de julho de 2021. Disponível em: <http://sistemas.saude.rj.gov.br/tabnetbd/dhx.exe?covid19/esus_sivep.def>

SILVA FILHO, Aloísio Machado da. Autocorrelação e correlação cruzada: teorias e aplicações. 2014. Disponível em: <<http://200.9.65.226/handle/fieb/766>>