



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

MARIANA SILVA MOTA

**METODOLOGIA PARA CONSTRUÇÃO DE TAXONOMIAS PARA SISTEMAS DE
RECOMENDAÇÃO: APRIMORANDO A PRECISÃO COM PLN E MÉTRICAS DE
AUDIÊNCIA**

FORTALEZA

2024

MARIANA SILVA MOTA

**METODOLOGIA PARA CONSTRUÇÃO DE TAXONOMIAS PARA SISTEMAS DE
RECOMENDAÇÃO: APRIMORANDO A PRECISÃO COM PLN E MÉTRICAS DE
AUDIÊNCIA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Ceará, como requisito para a obtenção do título de mestre em Ciência da Informação.

Área de concentração: Representação e Mediação da Informação e do Conhecimento.

Linha de Pesquisa: Informação, organização, comunicação e tecnologia.

Orientador: Prof. Dr. Osvaldo de Souza.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M871m Mota, Mariana Silva.

Metodologia para construção de taxonomias para sistemas de recomendação :
Aprimorando a precisão com PLN e métricas de audiência / Mariana Silva Mota. – 2024.
132 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Humanidades,
Programa de Pós-Graduação em Ciência da Informação, Fortaleza, 2024.
Orientação: Prof. Dr. Osvaldo de Souza.

1. taxonomia. 2. sistemas de recomendação. 3. experiência do usuário. I. Título.

CDD 020

MARIANA SILVA MOTA

**METODOLOGIA PARA CONSTRUÇÃO DE TAXONOMIAS PARA SISTEMAS DE
RECOMENDAÇÃO: APRIMORANDO A PRECISÃO COM PLN E MÉTRICAS DE
AUDIÊNCIA**

Aprovada em: 22/08/2024

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Ceará, como requisito para a obtenção do título de mestre em Ciência da Informação.

BANCA EXAMINADORA

Prof. Dr. Osvaldo de Souza (Orientador)
Universidade Federal do Ceará (UFC)

Profa. Dra. Silvana Vidotti
Universidade Estadual Paulista (UNESP)

Profa. Dra. Maria Giovanna Guedes Farias
Universidade Federal do Ceará (UFC) / suplente

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão aos meus pais, Zenete e Vicente, que sempre me incentivaram a buscar o mestrado, acompanharam-me em todo o processo e comemoraram comigo cada etapa, apoiando-me a nunca desistir.

À minha irmã, Maria Clara, que sempre acreditou no meu potencial e me incentivou a sair do meu quarto para momentos ao ar livre; frequentemente me pergunto o que seria da minha saúde mental sem você.

Agradeço também à minha tia, Nazete, que faleceu no início do mestrado e não pôde acompanhar meus primeiros meses de aula, mas que, mesmo na UTI, sorriu diversas vezes em resposta à minha conquista. E não posso deixar de mencionar que todos da minha família Silva e Mota tiveram um grande papel de apoio nesse processo. Obrigada.

Aos meus amigos, que constantemente me ouviram falar sobre o assunto e, em vez de pedir para parar, me acolheram em todos os momentos, me deram dicas, suporte e apoio, respeitaram meu espaço quando eu precisava me recolher para estudar, demonstrando que a verdadeira amizade nos acompanha em momentos bons e complicados. Agradeço a cada um de vocês.

Ao meu orientador, Osvaldo de Souza, que me acompanhou nessa jornada desde o início. Embora eu não esperasse que o mestrado fosse uma jornada linear, também não esperava tantas complicações pelo caminho. A vida acontece. Mesmo assim, você respeitou meu momento e me apoiou quando mais precisei. Sua fé no potencial desta ideia me fez acreditar ainda mais.

Agradeço a todos da coordenação e aos professores que me acompanharam até aqui, em especial, o professor Tadeu Feitosa e a secretária Veruska Maciel, que me deram suporte emocional, boas risadas e muitas conversas nos corredores e que foram essenciais nos dias em que parecia impossível conciliar tantas coisas.

Por fim, gostaria de trazer minha fé para encerrar os agradecimentos, com uma frase que um dia vi, mas não lembro exatamente onde: "Tudo que desejo é arriscado, mas lemanjá sempre me deu força e clareza para eu fazer uma passagem luminosa". Agradeço a lemanjá por me guiar até o encerramento deste ciclo. Água parada apodrece, e confio que ela me reserva novos rumos e que as ondas levarão as palavras deste trabalho para os lugares certos. Parei de temer a vida e de fugir de novas experiências porque ela sempre me possibilitou viver o impossível.

RESUMO

Este estudo tem como objetivo propor uma metodologia para a construção de taxonomias, utilizando uma abordagem metodológica em duas fases. A primeira fase, denominada Análise UX Integrada, combina métodos qualitativos e quantitativos para investigar as necessidades dos usuários e identificar tendências de mercado. A segunda fase, intitulada Metodologia para Taxonomias, concentra-se no desenvolvimento das etapas para a criação de taxonomias e na definição de seus cenários de aplicação em sistemas de recomendação, com foco na escalabilidade e eficiência do uso em tais sistemas. O estudo busca propor uma metodologia detalhada, delineando cada etapa do processo de construção das taxonomias. Especial atenção é dada à influência da experiência do usuário na estruturação da informação, com destaque para padrões de consumo que impactam diretamente o comércio eletrônico. O estudo também abrange os objetivos específicos de compreender o contexto do usuário por meio de técnicas de experiência do usuário (UX) e a análise de termos extraídos via processamento de linguagem natural (PLN), avaliando o desempenho desses termos em sistemas de recomendação. Este estudo faz uso de diversas ferramentas, incluindo questionários, bases de dados própria, análise quantitativa e qualitativa, além de estudos de mercado, permitindo uma análise abrangente dos contextos dos usuários e a subsequente modelagem da informação para a criação de taxonomias. Os resultados ressaltam a importância central do usuário no desenvolvimento de produtos e serviços, demonstrando que uma taxonomia bem estruturada facilita não apenas a categorização e o aprendizado automatizado, mas também o consumo eficiente de informações. A metodologia proposta revela-se de fácil implementação, contudo, apresenta grande potencial em termos de flexibilidade para diferentes cenários e precisão nas etapas de construção de taxonomias, evidenciando a relevância dos saberes da Ciência da Informação, como a taxonomia, no avanço de soluções tecnológicas e na otimização de sistemas de recomendação. Por fim, o estudo contribui para a validação da metodologia proposta, expandindo a compreensão sobre as interações entre a Ciência da Informação e a tecnologia, além de abrir novas oportunidades para pesquisas futuras nesse campo.

Palavras-chave: taxonomia; sistemas de recomendação; experiência do usuário.

ABSTRACT

This study aims to propose a methodology for building taxonomies using a two-phase methodological approach. The first phase, called Integrated UX Analysis, combines qualitative and quantitative methods to investigate user needs and identify market trends. The second phase, titled Taxonomy Methodology, focuses on developing the steps for creating taxonomies and defining their application scenarios in recommendation systems, with an emphasis on scalability and efficiency in such systems. The study seeks to propose a detailed methodology, outlining each step of the taxonomy-building process. Special attention is given to the influence of user experience on information structuring, with a focus on consumption patterns that directly impact e-commerce. The study also covers specific objectives such as understanding the user context through user experience (UX) techniques and the analysis of terms extracted via natural language processing (NLP), assessing the performance of these terms in recommendation systems. This study employs various tools, including surveys, proprietary databases, quantitative and qualitative analysis, as well as market studies, allowing for a comprehensive analysis of user contexts and the subsequent information modeling for taxonomy creation. The results highlight the central importance of the user in the development of products and services, demonstrating that a well-structured taxonomy not only facilitates categorization and machine learning but also enables efficient information consumption. The proposed methodology proves to be easy to implement while offering great potential in terms of flexibility for different scenarios and precision in the taxonomy-building steps, emphasizing the relevance of Information Science knowledge, such as taxonomy, in advancing technological solutions and optimizing recommendation systems. Finally, the study contributes to the validation of the proposed methodology, expanding the understanding of the interactions between Information Science and technology, while also opening new opportunities for future research in this field.

Keywords: taxonomy; recommendation system; user experience.

LISTA DE FIGURAS

Figura 1 - Fluxo das fases e etapas.....	65
Figura 2 - UX Persona - Usuário de sistema de recomendação e-commerce.....	90
Figura 3 – Taxonomia de interesse do Pinterest.....	112

LISTA DE GRÁFICOS

Gráfico 1 - Demografia regional dos entrevistados.....	17
Gráfico 2 - Diferença entre classificação e clusterização.....	50
Gráfico 3 – Dispositivos utilizado na compra dos brasileiros de acordo com Abcomm.....	80
Gráfico 4 - Dispositivo de preferência para fazer compras online.....	80
Gráfico 5 – Perfil dos Compradores Online (Faturamento) - Categorias de consumo.....	82
Gráfico 6 - Distribuição das categorias de consumo dos usuários entrevistados...	83
Gráfico 7 – Sites de preferência para fazer compras online.....	85
Gráfico 8 - Frequência da navegação através das recomendações dos produtos.	88

LISTA DE PLANILHAS

Planilha 1 - Termos e respectivas frequências para ar-condicionado.....	98
Planilha 2 - Termos e respectivas frequências para conectividade.....	100
Planilha 3 – Termos e respectivas frequências para contexto de uso – ar- condicionado.....	102
Planilha 4 - Termos e respectivas frequências para utensílios para iluminação....	104
Planilha 5 – Tabela de dados para categoria beleza.....	113

LISTA DE QUADROS

Quadro 1 - Características gerais das taxonomias: tradicional, de audiência e PLN.....	45
Quadro 2 - Etapas e respectivas definições.....	75
Quadro 3 – Entradas e saídas de cada etapa.....	76
Quadro 4 - Nomenclatura das taxonomias de conteúdo.....	95
Quadro 5 – Definição de tamanho das taxonomias de conteúdo.....	95

LISTA DE SIGLAS E ABREVIATURAS

Abcomm	Associação Brasileira de Comércio Eletrônico
ADTCT	Construção Automática de Taxonomia de Domínio a Partir de Texto
IA	Inteligência Artificial
NaLUE	Natural Language Understanding Uncertainty Evaluation
OI	Organização da Informação
PAM	Partitioning Around Medoids
PLN	Processamento de Linguagem Natural
TI	Tecnologia da Informação
UX	User Experience

SUMÁRIO

1 INTRODUÇÃO	12
2 METODOLOGIA	16
2.1 Fase 1: análise UX integrada	17
2.2 Fase 2: metodologia para taxonomias	19
2.2.1 Desenvolvimento de Taxonomias	20
3 A INFLUÊNCIA DA WEB COLABORATIVA NA ORGANIZAÇÃO DA INFORMAÇÃO ...	22
4 O AMBIENTE DIGITAL E A CENTRALIZAÇÃO DO USUÁRIO	29
4.1 A experiência do usuário	31
5 O VALOR DA TAXONOMIA	36
5.1 Comparação entre Taxonomia tradicional e Taxonomia de audiência	38
5.2 Comparação entre taxonomia tradicional e taxonomia PLN	41
5.3 Prós e contras de uma taxonomia para um sistema recomendação	44
6 CLUSTERIZAÇÃO E RECOMENDAÇÃO DE CONTEÚDO	48
6.1 Tipos de organização necessários no processo de clusterização	49
7 SISTEMAS DE RECOMENDAÇÃO	55
7.1 Exemplos de uso	56
7.2 Prós e contras dos sistemas de recomendação	60
8 DESENVOLVIMENTO DA METODOLOGIA DE CONSTRUÇÃO DE TAXONOMIA	63
8.1 Uma metodologia de construção de taxonomia	63
8.1.1 Fase 1: Análise UX Integrada.....	70
8.1.2 Fase 2: Metodologia para Taxonomias	72
8.2 Experimentação e Validação	73
8.2.1 Fase 1: Análise UX Integrada.....	78
8.2.2 Fase 2: Metodologia para Taxonomias	91
9 CONSIDERAÇÕES FINAIS	118
REFERÊNCIAS	120
APÊNDICE A - TAXONOMIA DE CONTEÚDO PARA ELETRODOMÉSTICOS – AR- CONDICIONADO	126
APÊNDICE B - TAXONOMIA DE CONTEÚDO - CONTEXTO DE USO EM AR- CONDICIONADO	127
APÊNDICE C - TAXONOMIA DE CONTEÚDO - UTENSÍLIOS PARA ILUMINAÇÃO	128
APÊNDICE D - TAXONOMIA DE AUDIÊNCIA – BELEZA	129
APÊNDICE D – ROTEIRO DA ENTREVISTA COM OS USUÁRIOS	130

1 INTRODUÇÃO

Há diversas possibilidades a serem abordadas no âmbito da tecnologia, cuja abrangência engloba o processamento de linguagem natural (PLN), o aprendizado de máquina, a mineração de dados, *Big Data*, a representação do conhecimento entre outros.

Uma questão fundamental que impulsionou a problemática desta pesquisa está relacionada ao desenvolvimento de uma estrutura de taxonomia escalável e capaz de aprimorar sistemas de recomendação. Em 2021, uma proposta de trabalho chamou a atenção por envolver a expertise necessária para construir taxonomias voltadas para sistemas de recomendação e treinamento em aprendizado de máquina para uma grande empresa de *streaming* com alcance global.

Esse aspecto gerou interesse sobre como o processo de desenvolvimento das taxonomias seria conduzido. Surgiram questionamentos sobre a existência de uma metodologia específica, os trâmites utilizados e de que forma a taxonomia poderia potencializar o processo.

Além disso, esse interesse despertou o anseio por uma metodologia que facilitasse o trabalho cotidiano na área de taxonomia aplicada à tecnologia. Esses questionamentos e inquietações culminaram no tema desta pesquisa, que tem como objetivo geral discutir e apresentar uma metodologia para o desenvolvimento de estruturas taxonômicas, com foco na taxonomia PLN e na taxonomia de audiência, cujas aplicações voltadas para a disseminação da informação visam aprimorar os sistemas de recomendação.

O desenvolvimento de metodologia de criação da taxonomia é importante para facilitar a escalabilidade da estrutura taxonômica - conceito este discutido mais à frente, no capítulo referente à clusterização - e viabilizar a organização eficaz de uma quantidade crescente de informações, bem como a geração de conhecimento constante dentro das bases de dados.

A escalabilidade desempenha um papel crucial no impacto sobre os negócios, na organização de informações em *Big Data* e na recuperação de dados. A melhor forma de viabilizar essa escalabilidade em diversos sistemas é por meio de uma ou mais estruturas taxonômicas.

Também há intenção em lançar luz sobre o campo da inteligência artificial sob a ótica dos estudos em Ciência da Informação, no sentido de investigar o

comportamento da organização informacional da estrutura taxonômica e respectivo impacto no sistema de recomendação. Desta forma, esta pesquisa emprega a análise quantitativa e o estudo de caso como procedimentos metodológicos, sendo estes exploratórios e quantitativos, fundamentado na dinâmica da pesquisa-ação com uma fase de pré-análise, identificação de termos, exploração do material e análise e desenvolvimento.

Nesse contexto, o primeiro objetivo específico é compreender o perfil do usuário e como ele orienta a modelagem da informação. A importância desse objetivo reside no fato de que, ao traçar o perfil do usuário, haverá uma influência significativa no desenvolvimento da taxonomia.

O segundo objetivo específico é analisar os termos extraídos do processamento de linguagem natural dentro de uma taxonomia, destacando seu impacto nos resultados de recomendação.

Para atingir os objetivos propostos, este trabalho está organizado em seis capítulos, da seguinte maneira:

O Capítulo 1 apresenta a introdução do estudo, oferecendo um resumo do conteúdo de cada capítulo do trabalho.

O Capítulo 2 aborda a metodologia utilizada para conduzir o processo de pesquisa, seguido pela descrição de como ocorre a coleta de dados relevantes.

No Capítulo 3, discute-se a influência da web colaborativa na organização da informação, o foco é fornecer um entendimento de um recorte histórico da Web 2.0 e sua centralização do usuário como criador e consumidor de conteúdo. Este capítulo é essencial para definir o cenário no qual sobressai a importância do estudo e desenvolvimento de produtos centrados no usuário. Também prepara o leitor para compreender o mundo centrado no usuário, observado a partir da mudança dos parâmetros da internet, na qual os usuários se tornaram protagonistas, foi um momento crucial na história da tecnologia digital. Como resultado, tornou-se essencial que quase todos os produtos digitais sejam projetados em direção à experiência positiva do usuário.

O capítulo 4 apresenta o mundo centrado no usuário. São abordados os ensinamentos, conteúdos e propostas do trabalho de Don Norman, que destacam a importância de projetar produtos e sistemas que sejam intuitivos, amigáveis e melhorem a experiência geral do usuário final, possibilitando clareza sobre a natureza

evolutiva da experiência do usuário e enfatizando sua crescente relevância no domínio do consumo online.

A importância desse capítulo é elucidar sobre como a evolução da experiência do usuário e como o rápido crescimento do comércio eletrônico, mídia social e outras plataformas digitais, tornou-se um fator crucial que determina o sucesso ou o fracasso de um produto ou serviço. Compreender e implementar os princípios descritos por Norman pode impactar significativamente a eficácia e a usabilidade das interfaces digitais, melhorando a satisfação do usuário e impulsionando o crescimento dos negócios.

O capítulo 5 da estrutura teórica discute a taxonomia, explora o conceito de taxonomia, seu valor inerente e traça comparações entre a taxonomia tradicional e a taxonomia de audiência, bem como as de processamento de linguagem natural.

Ao final desse capítulo, apresenta-se uma avaliação das vantagens e desvantagens associadas à utilização da taxonomia.

Por sua vez, o capítulo 6 da pesquisa aborda a clusterização e recomendação de conteúdo; tem como foco a abordagem do agrupamento de usuários, apresentando como o agrupamento de usuários pode ser efetivamente implementado em uma instituição, destacando os vários fatores que influenciam o processo de agrupamento. Além disso, este capítulo oferece uma análise abrangente do processo de agrupamento, esclarecendo os métodos e técnicas empregados para agrupar usuários com base em suas semelhanças e características comuns. Este capítulo permite uma investigação e análise aprofundadas dos recursos, capacidades e limitações da ferramenta escolhida.

O capítulo 7 traz para a discussão os sistemas de recomendação, os quais giram em torno do impacto significativo da influência do usuário no consumo e análise de dados e a recomendação de informação e produtos. Este capítulo serve como uma ponte crucial, conectando os fundamentos teóricos apresentados nos capítulos anteriores à aplicação prática da análise de dados em cenários do mundo real.

Ademais, visa esclarecer o relacionamento entre usuários e dados, destacando como o comportamento, as preferências e os vieses do usuário podem moldar os resultados da análise de dados, demonstrando a exigência de sistemas sofisticados de recomendação para fomentar uma maior interação por parte do usuário no consumo do produto/informação.

Por fim, o capítulo 8 abrange todo o desenvolvimento do projeto e apresenta a proposta de uma metodologia para a criação de taxonomias. As considerações finais, encerrando as ideias discutidas, são apresentadas no capítulo 9. O trabalho é concluído com as referências e apêndices, dispostos ao final.

2 METODOLOGIA

A metodologia empregada no presente trabalho consistiu em uma pesquisa exploratória, de caráter bibliográfico, na qual foram identificados os teóricos e textos necessários ao primeiro contato com o tema da pesquisa, cujas análises e discussões sobre estes textos encontram-se distribuídas entre os capítulos 3 até o capítulo 7. Foram encontrados também textos fortemente relacionados aos objetivos do presente trabalho, os quais são discutidos mais adiante no capítulo 8. Quanto aos procedimentos de análise, esses foram qualitativos na parte exploratória da pesquisa; e quali-quantitativos na etapa de desenvolvimento da metodologia proposta, apresentada no capítulo 8, a qual contou com uma coleta de dados através de formulários e e-mails, visando obter insumos para uma validação da metodologia proposta em um estudo de caso simples.

Os textos que propunham etapas ou roteiros para criação de taxonomias foram examinados em detalhes, o que permitiu evidenciar as características, pontos fortes e fracos de cada um dos métodos examinados - tais discussões ocorrem no capítulo 8.

Na etapa de desenvolvimento da pesquisa, adotou-se uma abordagem combinatória em duas fases. A primeira fase, nomeada Fase 1: Análise UX Integrada, na qual adota-se uma abordagem qualitativa e quantitativa, focada em entender como a necessidade do usuário e as tendências do mercado influenciam as necessidades a serem consideradas na construção de taxonomias. Tais achados eram vitais para viabilizar a segunda fase, nomeada Fase 2: Metodologia para Taxonomias. Esta segunda fase é direcionada ao desenvolvimento da metodologia de criação da taxonomia e aos consequentes cenários de uso em sistemas de recomendação.

No sentido de clarificar o andamento da pesquisa em si e a construção intelectual da metodologia resultante da pesquisa, optou-se por apresentar os procedimentos metodológicos de ambas as fases 1 e 2 da etapa de desenvolvimento nesta seção.

A razão para tal foi diminuir a possível sobreposição de conceitos e momentos da pesquisa, visto que a metodologia proposta também será estruturada em fases e etapas, como apresentado no capítulo 8. Assim, deve-se compreender a separação entre o trabalho realizado na pesquisa e a estruturação da metodologia proposta, que em momentos tem nomes e propósitos semelhantes. As próximas seções, portanto,

esclarecem os procedimentos metodológicos da pesquisa no processo de construção de uma proposta da metodologia de criação de taxonomia.

2.1 Fase 1: análise UX integrada

Tem como propósito analisar o cenário e experiência do usuário e do mercado para otimizar os resultados do sistema de recomendação.

A análise de experiência do usuário foi conduzida em três momentos distintos: a) Descoberta (*Discovery*); b) Referência (*Benchmark*); e c) Refinamento.

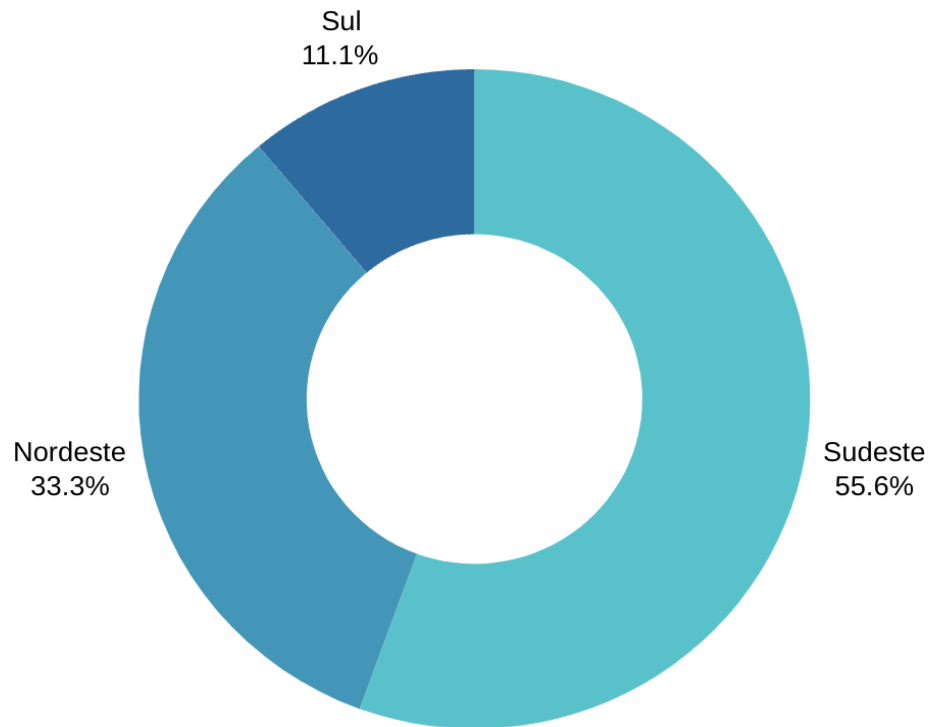
Na descoberta para a coleta de dados foi realizado um questionário utilizando o Google Forms, com o objetivo de alcançar um maior número de usuários que utilizam quaisquer tipos de sites de compras online e consomem produtos recomendados.

O propósito foi identificar as facilidades e desafios que esses usuários enfrentam ao utilizar os sistemas de recomendação em sites de e-commerce. O formulário foi divulgado no perfil pessoal aberto da rede social X (antigo Twitter) para ampliar o alcance de possíveis participantes. Para as pessoas que demonstraram interesse, foi enviado um e-mail formalizando a participação, explicando o objetivo da pesquisa e a importância de sua colaboração. A coleta dos dados ocorreu no recorte cronológico de 20/02/2024 até 11/05/2024, e resultou em uma amostra da seguinte forma: nove pessoas se dispuseram a ajudar na pesquisa. A faixa etária dos participantes é entre 24 e 38 anos.

Tais dados foram analisados durante o momento de referência, o que revelou que o público participante está dividido geograficamente no Brasil, como se vê no Gráfico 1, da seguinte forma: três pessoas do Nordeste, uma do Sul e cinco da região Sudeste, destacando-se o estado de São Paulo com o maior número de participantes. A amostra também está dividida em 44,4% masculino e 55,6% feminino.

No momento de análise de *benchmarking*, o esforço foi direcionado em compreender a disposição das informações e categorizações feitas no mercado, pois isso fornece uma base sólida para a estrutura taxonômica e ajuda a compreender melhor as categorias a serem exploradas e a se inspirar no material existente. Como um dos resultados do *benchmarking*, elegeu-se os tipos de categorização utilizados por atores relevantes do segmento comercial que tem ampla aceitação no mercado e que tenham sido mencionados pelos participantes da pesquisa.

Gráfico 1 - Demografia regional dos entrevistados



Fonte: Elaborado pela autora (2024).

Nesse sentido, baseado nos resultados obtidos, foram investigadas as categorizações da Shopee, Amazon e AliExpress. Conforme destacado por Schrepp, Thomaschewski e Hinderks (2017), o *benchmarking* fornece um entendimento inestimável sobre os pontos fortes e fracos da experiência do usuário em um determinado cenário ou produto, permitindo que os pesquisadores identifiquem áreas para melhoria e tomem decisões baseadas em dados destinadas a aprimorar o desempenho do usuário.

Ao estabelecer referências claras e avaliar os dados coletados na descoberta, as organizações podem obter uma compreensão abrangente de como seus produtos ou serviços atendem às expectativas do usuário e aos padrões de usabilidade predominantes

No último momento, que é o de refinamento, os dados obtidos por meio das entrevistas e das referências são novamente analisados e cruzados. Essa nova

análise permite a identificação de padrões que só se tornaram evidentes após a conclusão das etapas anteriores.

Com base nessas informações, foi possível refinar e estruturar os dados coletados, criando assim a *persona* que irá orientar a aplicação na metodologia de criação das estruturas taxonômicas. Uma *persona* é uma abstração e simulação de representação geral de um possível usuário baseada em uma pesquisa. O uso deste mecanismo ajuda a entender melhor os desejos, frustrações, necessidades e comportamentos do usuário, ajudando a nortear o desenvolvimento de soluções mais direcionadas e objetivas.

As *personas* são representadas por uma imagem que inclui uma foto e um nome fictício do usuário, dados demográficos, contexto em que o usuário está inserido, necessidades principais, frustrações, plataformas e mídias sociais mais utilizadas e dispositivos principais. Além disso, há uma citação que resume bem o sentimento dos usuários entrevistados na pesquisa.

Os dados coletados e utilizados constituem o conjunto mínimo para formar uma *persona* em qualquer trabalho, incluindo este, mas outras informações podem ser adicionadas ou excluídas, dependendo do escopo e dos objetivos de cada projeto.

Foram dedicadas quatro semanas e meia na etapa de experiência do usuário, duas semanas de descoberta, uma semana de referência e uma semana e meia para refinamento.

O roteiro do questionário aplicado aos usuários consta no apêndice E.

2.2 Fase 2: metodologia para taxonomias

A segunda fase tem como propósito prover elementos necessários para a investigação de uma nova abordagem na construção de estruturas taxonômicas para sistemas de recomendação. Os trabalhos foram divididos nos seguintes esforços: a) análise dos termos, b) desenvolvimento de taxonomias. Ambos são explicados nas próximas seções.

A Análise de Termos é parte integrante da Fase 2 e inicialmente estava sendo realizada utilizando-se a base de dados gratuita NaLUE do Google, a qual infelizmente durante a pesquisa iniciada em 2022, tornou-se inacessível. Por isso, foi necessário recorrer a exemplos da minha experiência como taxonomista ao longo de quatro anos, no período entre 2020 e 2024, na qual há o registro contínuo proveniente das

atividades profissionais, tabulados e organizados com diversos termos e comentários que me permitem usá-la como guia de memória para orientar a fase de termos e desenvolvimento da taxonomia. No entanto, não posso compartilhar a planilha e os termos na íntegra, pois assinei um acordo de sigilo de 10 anos, o uso de fragmentos necessários para esta pesquisa foi autorizado pelo empregador.

Portanto, os termos que serão usados neste trabalho são baseados nessa "base de dados" pessoal de 20 mil termos. A abordagem adotada foi: visitar os sites AliExpress, Shopee e Amazon, selecionar comentários e coletar termos recorrentes entre eles para criar uma base de dados fictícia. Ressalto que o site AliExpress já fornece os principais termos mais comentados de qualquer produto, o que facilitou a pesquisa e a coleta de dados. A escolha desses sites deu-se em função de terem sido mencionados na coleta de dados junto aos participantes da pesquisa.

Assim, utilizando essa base de dados, foi possível exemplificar o desenvolvimento da taxonomia de forma adequada, garantindo a qualidade e a relevância dos termos para a construção da estrutura taxonômica.

2.2.1 Desenvolvimento de Taxonomias

Quanto ao desenvolvimento de taxonomias, nesta pesquisa foram desenvolvidas quatro taxonomias simultaneamente para testar a proposta da metodologia e sua aplicabilidade universal em diferentes contextos. Essas taxonomias são relacionadas aos seguintes domínios: a) Eletrodomésticos, b) Utensílios para Casa, c) Contexto de Uso e d) Beleza. Os três primeiros são voltados para a semântica, formando a taxonomia de conteúdo; enquanto a última, a taxonomia de Beleza, é voltada para os segmentos de mercado, formando uma taxonomia de audiência.

Considerando o escopo e o prazo desta pesquisa, as taxonomias desenvolvidas não foram testadas com dados em sistemas computacionais de recomendação para experimentar a metodologia proposta e seu respectivo impacto sistêmico. Para isso, seria necessário construir um cenário e realizar uma exploração minuciosa dos dados, estudar e acompanhar os impactos no sistema computacional de recomendação, verificando se há evolução após a implementação da taxonomia, e comparar dois cenários: um antes, sem a metodologia de criação de taxonomia, e outro depois, com a metodologia aplicada. A construção do sistema computacional e

todo o cenário de teste necessários seriam extensos e envolveriam uma equipe de trabalho, motivo pelo qual não foi incluída essa verificação como escopo do mestrado, ficando relacionado como uma das possibilidades para produções futuras.

Deste modo, neste estudo foram incluídos apenas os esforços relativos a uma análise das possibilidades de aplicação da metodologia proposta, sendo a validação da metodologia realizada teoricamente com a produção de taxonomias de exemplo, detalhadas no capítulo 8.

3 A INFLUÊNCIA DA WEB COLABORATIVA NA ORGANIZAÇÃO DA INFORMAÇÃO

O mundo, de um modo geral, atrai-nos por sua beleza, complexidade e diversidade, de modo que tudo isso ocorre simultaneamente. Na maior parte do tempo, estamos imersos e atraídos a uma quantidade demasiada de elementos diversos que o nosso ambiente pode nos oferecer.

Atualmente, não existe somente a informação de símbolos do cotidiano. Houve um alto impacto desde a inovação de Gutemberg, que foi responsável por acelerar a produção e disseminação do conhecimento registrado, mas que desencadeou alguns problemas na modernidade, como o excesso informacional (Sales *et al.*, 2016).

No cenário atual, com a internet, conectamo-nos a universos distintos e simultâneos. A revolução promovida pela internet permitiu a expansão da interconectividade em diversas esferas do conhecimento, abrangendo desde a literatura, periódicos até a cinematografia e finanças, como também popularizou novas ferramentas de acesso às informações, como leitores eletrônicos, mídias sociais e transações financeiras digitais.

Essa expansão impacta diversos setores econômicos, políticas globais e até mesmo a vida cotidiana da sociedade. Exemplos disso incluem a transição de uma carta escrita à mão para mensagens em tempo real. Antes, os comentários soltos na rua agora se manifestam como comentários nas redes sociais, onde a quantidade de curtidas amplia o alcance da opinião dos usuários a níveis nunca imaginados.

Essa expansão vai além das interações sociais. O que antes era um simples alarme em um relógio analógico, agora se tornou alarme do celular e até mesmo sistemas IA ativados por voz em casas inteligentes, como o Alexa da Amazon.

Atualmente, a IA está sendo utilizada para uma ampla gama de atividades, incluindo diagnóstico médico, plataformas de negociação eletrônica, controle de robôs e sensoriamento remoto. Tem sido usado para desenvolver e promover vários campos e indústrias, incluindo finanças, saúde, educação, transporte e robótica (Kayid, 2020, p. 3, tradução nossa)

A difusão e o uso das tecnologias mencionadas por Kayid (2020) abrangem diversos aspectos do cotidiano social e seu impacto nas esferas econômica, financeira, de transporte e educação. Esse cenário destaca um movimento de

enraizamento social profundo da Inteligência Artificial. Como sociedade, estamos testemunhando uma diversidade sociocultural em que sistemas inteligentes podem oferecer respostas imediatas para uma variedade de tarefas. Essas dinâmicas coletivas e participativas estão cada vez mais integradas à Inteligência Artificial.

Lévy (2009) alude que os laços sociais são anteriores às influências tecnológicas, com o processo de virtualização apenas amplificando a nossa propensão para a interconectividade, revivendo assim os constituintes fundamentais que moldam a nossa realidade experiencial. O domínio virtual não só ampliou o espectro de mecanismos colaborativos, como também simboliza a base do paradigma da Web 2.0.

O termo "Web 2.0" surgiu na sequência de um discurso colaborativo entre Tim O'Reilly (2005) e Media Live International delineando uma mudança paradigmática que reposicionou o usuário como ponto focal da geração de conteúdos, o protagonista. A eficácia desse protagonismo online depende significativamente da participação colaborativa dos usuários, uma vez que estabelece o papel de cocriação de conteúdo.

Os atributos distintos da Web 2.0 abrangem o desenvolvimento de aplicações que funcionam nos modos *online* e *offline*, o desenvolvimento de comunidades virtuais ativas, o aumento dos esforços de comércio eletrônico, a adoção de *tags* que exercem significado semântico em textos para facilitar a recuperação de informações e a facilitação do acesso recíproco e manipulação da informação (Nascimento; Quintão, 2011).

Exemplos de sites relevantes no início da era do conteúdo colaborativo utilizando predominantemente as *tags* incluem MySpace e Tumblr. Estes sites simbolizam uma fase crucial da Web 2.0, caracterizada pela expansão, estruturação e envolvimento dos usuários, oferecendo perspectivas sem precedentes para participantes ativos.

Um princípio fundamental para essa participação ativa era simples: quanto mais ampla for a base de utilizadores envolvidos, maior será o potencial de alcance. Consequentemente, as distâncias geográficas foram transcendidas, permitindo a interligação de indivíduos em diversos locais e promovendo um sentido de comunidade virtual coesa e interligada.

As *tags*, além de servirem como ferramentas de indexação para recuperação de informações, adquirem uma conotação aumentada, representando a voz individual,

as ideologias, a afetividade e os atributos culturais de cada participante da rede. À medida que o alcance destas plataformas expandiu, os criadores de conteúdos assumiram proeminência, com a sua visibilidade independentemente do tamanho dos seus seguidores e da influência que exerciam sobre os indivíduos e as suas predileções de consumo.

De fato, essas mesmas *tags* começaram a transcender a afiliação à comunidade, assumindo importância em termos de alcance, fama e ganho financeiro. As preferências do usuário, incluindo estilo de vida, hábitos alimentares, escolhas de moda, compromissos sociais, padrões de gastos, preferências musicais, bem como opiniões sobre figuras públicas gerou uma riqueza de informações sobre o consumidor.

Essa gama de informações produzidas e de fácil acesso permitiu os usuários influenciarem e serem influenciados, de modo a construir uma diferente cadeia intrincada de participantes ativos, em que o mundo virtual e real se fundiu, oferecendo um retrato detalhado do perfil de cada usuário. Essa perspectiva é aludida por Lima e Alvares (2012):

A informação obtida por um indivíduo, para se transformar em conhecimento, dialoga com a sua cultura, seus valores e princípios, seu modo de ser e sua maneira de ver e compreender o mundo. O conhecimento, nesse caso, é subjetivo (inerente ao sujeito), mas ao mesmo tempo social, pois o ser humano interage com o mundo que o circunda, modificando-o e sendo por ele modificado. Nem toda informação existente em um documento vai se transformar em conhecimento, pois quem aprende precisa ter os elementos fundamentais para a decodificação da informação, ou seja, fazer a correlação dessa informação com as estruturas mentais e conhecimentos correlatos mínimos que possibilitarão o entendimento e, se for o caso, a geração de novos conhecimentos (Lima; Alvares, 2012, p. 25).

Esta última citação aponta sobre a transformação da informação em conhecimento e como essa transformação ocorre através de diversos elementos, como valores e cultura. Isso é um movimento constante na Web 2.0, como as *tags* são a representação do usuário para um certo tipo de informação, com ele irá seu viés, cultura, valores, crenças e entre outros; possibilitando assim que pessoas que sejam do seu círculo social, ou que possuem a mesma linguagem, gosto, compartilhem dos mesmos valores sejam capazes de obter acesso aos documentos que esse usuário disponibilizou online.

Esse movimento de representar e compartilhar a partir do viés de cada usuário parece simples, mas foi um grande catalizador do compartilhamento e gerou conceitos

- hoje, conhecido como folksonomia (Wal, 2007), a forma relacional de compartilhar conteúdos - e proporcionou aos usuários a melhor experiência possível com um novo senso de comunidade e relação com a informação, além de ser preciso uma nova forma de desenvolver uma Organização da Informação (OI).

O objetivo da organização informacional é impactar positivamente a utilidade e usabilidade dos produtos produzidos por grandes empresas de tecnologia. Conseqüentemente, podemos inferir que a disciplina de organização da informação desempenha um papel crucial no consumo informacional. Segundo Kielgast e Hubbard (1997), a Organização da Informação facilita o acesso eficiente à informação e agrega valor ao processo como um todo.

Um dos principais benefícios da OI é a economia de tempo durante a recuperação de informações. Além disso, a organização da informação aumenta nossa compreensão do mundo, facilita a associação de significados e auxilia na definição de várias entidades.

Para cumprir os objetivos da Organização da Informação, os objetos informacionais requerem descrições físicas e de conteúdo, sendo a aquisição de conhecimento o objetivo final, conforme articulado por Brascher e Café (2008):

Para que os objetivos da OI sejam alcançados, é preciso realizar a descrição física e de conteúdo dos objetos informacionais. A descrição de conteúdo tem por objeto o primeiro dos três elementos da informação propostos por Fogl – o conhecimento. A descrição física, por sua vez, direciona-se ao terceiro elemento - o suporte da informação. O segundo elemento – a linguagem – permeia os dois tipos de descrição (Brascher; Café, 2008, p. 5).

Em conjunto com o consumo massivo de informações, outro fator crucial que impacta significativamente os usuários é a assimilação de conteúdos categorizados por meio de recomendações do sistema. Sobre esse consumo massivo e correspondente volume de oferta de conteúdos na web, Souza e Fernandes (2020, p. 146) discorrem:

Há, portanto, um aumento do tempo necessário para que um potencial usuário de uma informação a encontre; obviamente isso é proporcional ao volume de informação disponível. Acreditamos, contudo, que é possível alterar essa relação com o tempo através do uso de tecnologia que minimize o volume de informações a serem apreendidas, quer seja pela diminuição do volume de texto a ser lido, ou pela alteração do tipo de código usado no suporte. Se um determinado documento puder ser transformado em outro tipo de documento, todavia preservando-se o valor semântico, poderíamos ter, por exemplo, uma condensação dos textos, criando versões de denso valor

semântico que requereriam menor tempo para seu consumo (Souza; Fernandes, 2020, p. 146).

Se por um lado há mais oferta, que é uma característica marcante do ecossistema informacional, ocorre também o fenômeno no qual passa a ser necessário mais tempo e recursos para que essa oferta possa ser percebida e analisada pelo seu potencial interessado.

Este fenômeno trouxe mudanças consideráveis na forma como a informação é apresentada e consumida. Elevou a proeminência de uma ampla gama de assuntos, abrangendo a proliferação de conteúdo gerado pelo usuário, as opiniões e perspectivas agora desempenham um papel significativo na formação do discurso público e na influência de tendências.

O impacto dos utilizadores de plataformas de criação e publicação de textos na web, na produção e distribuição de informação online é inegável. Essa mudança de paradigma colocou o usuário no centro do cenário digital. Plataformas e serviços são cada vez mais projetados para atender as necessidades e preferências individuais.

Um exemplo de serviço projetado para atender necessidades individuais são os *chatbots*, esses programas conversacionais multifacetados estão disponíveis para fornecer diversos tipos assistência personalizada, desde acionar por comando de voz uma máquina para lavar uma roupa até solicitar uma música.

Os chatbots podem imitar conversas humanas e entreter os usuários, mas não foram criados apenas para isso. Eles são úteis em aplicações como educação, recuperação de informações, negócios e comércio eletrônico. Eles se tornaram tão populares porque os chatbots também oferecem muitas vantagens para usuários e desenvolvedores. A maioria das implementações é independente de plataforma e está disponível instantaneamente para os usuários, sem a necessidade de instalações (Adamopoulou; Moussiades, 2020, tradução nossa).

Além dos *chatbots*, os serviços de *streaming* empregam sistemas complexos de recomendação de filmes e áudio para atender a personalização de conteúdo, os sistemas robustos de recomendação de produtos em sites de comércio eletrônico reconhecem as preferências do usuário e recomendam produtos de acordo com o perfil do usuário correspondente.

Este encadeamento de informações personalizada visa melhorar a experiência do usuário e estabelecer uma dependência mais forte nas plataformas. O

consumo de informação não é mais determinado apenas pelos meios de comunicação tradicionais ou abordagens de cima para baixo. Os usuários ganharam influência significativa por meio de seu envolvimento e feedback.

As recomendações do sistema, geradas pelos próprios usuários e facilitadas pelos sistemas de comércio eletrônico, transformaram a forma como as informações são organizadas e distribuídas. Essa abordagem centrada no usuário, alimentada por aprendizado de máquina e personalização, deu início a uma nova era de experiências online, onde os indivíduos têm mais controle e influência sobre o conteúdo que consomem, levando a um ambiente digital mais personalizado, imersivo e poderoso.

Aproveitando o tópico sobre personalização, é possível iniciar um tema mais visionário, a Web descentralizada, também conhecida como Web Semântica ou Web 3.0, a personalização é um dos fatores que representa a mudança da 2.0 para a 3.0.

O terceiro estágio da web, conhecido como Web 3.0, ainda é emergente e se caracteriza por uma mudança para sistemas mais inteligentes, descentralizados e interconectados, permitindo interações mais seguras, eficientes e personalizadas. A Web 3.0 baseia-se em tecnologias como blockchain, inteligência artificial e a web semântica, com foco em permitir interações descentralizadas e ponto a ponto, mais transparentes, confiáveis e resilientes. Plataformas Web 3.0 visam fornecer maior controle e propriedade dos dados aos usuários, facilitando transações mais seguras e eficientes. No geral, a evolução da web, da Web 1.0 para a Web 3.0, reflete uma progressão em direção a sistemas mais inteligentes, interativos e descentralizados, que permitem processos mais personalizados e interações seguras e eficientes entre usuários e máquinas. Cada estágio da web foi construído sobre os fundamentos do anterior, levando a sistemas cada vez mais complexos e sofisticados, com potencial para transformar diversas indústrias e setores. (Bharadiya, 2023, tradução nossa).

A diferença entre as webs 1.0, 2.0 e 3.0 está no fato de que a 1.0 disponibilizou informações online sem a participação ativa dos usuários, apenas permitindo o consumo dessas informações, enquanto a 2.0 trouxe interatividade e participação dos usuários na criação e consumo de conteúdo. A 3.0 representa a evolução desses aspectos, com sistemas cada vez mais inteligentes e interativos.

Bharadiya (2023) apresenta uma visão futurista da internet 3.0, destacando sua interatividade, segurança, personalização e inteligência aprimorada, impulsionada por importantes tecnologias, como realidade virtual, web semântica, inteligência artificial e *blockchains*.

Uma característica crucial mencionada é a capacidade de personalização que a web 3.0 oferece, tornando os sistemas ainda mais sofisticados.

Embora o autor não tenha abordado explicitamente a experiência do usuário em seu texto, gostaria de enfatizar a importância da centralização no usuário como elemento fundamental para a consolidação da web 3.0 e como um fator determinante para a personalização. Analisar as necessidades do usuário e buscar atendê-las proporciona um alto grau de personalização e sofisticação aos diversos sistemas, sendo, portanto, um aspecto crucial na transição para a web 3.0.

4 O AMBIENTE DIGITAL E A CENTRALIZAÇÃO DO USUÁRIO

No ambiente digital, os usuários estão inseridos em um universo em que o processo evolutivo das plataformas é constante. Essas plataformas buscam não só oferecer uma interface mais bonita, com mais opções de cores e resoluções de imagem e vídeo, mas também melhorar a experiência de uso em cada uma delas, como por exemplo, apostando em personalização de conteúdos com elaborados algoritmos de recomendação. A revista Forbes (Tilia, 2024) fez uma matéria recente e trouxe um caso interessante para representar esse exemplo, a rede social Tiktok, que está cada vez mais se tornando uma importante ferramenta de busca, e dependendo do público-alvo, que no caso é a geração Z, está superando o famoso e tradicional Google.

Um estudo feito por Southern (2024) revela que 64% da geração Z (pessoas que nasceram após 1996) prefere utilizar o Tiktok como uma ferramenta de pesquisa, e 10% desse público também utiliza o Chat GPT.

A pesquisa mostra que 40% dos consumidores agora usam o TikTok para pesquisar coisas de que precisam ou que têm interesse em aprender. Esta tendência prevalece entre os jovens, com 64% da Geração Z (aqueles nascidos depois de 1996) e 49% dos millennials (nascidos entre 1981-1996) recorrendo ao TikTok como ferramenta de pesquisa. A plataforma está sendo usada para pesquisar todos os tipos de tópicos – desde receitas culinárias, que a Geração Z pesquisa 29% mais do que a geração Y, até música, ideias DIY, moda e muito mais. Quase 10% dos usuários da Geração Z agora preferem usar o TikTok em vez de mecanismos de busca estabelecidos como o Google ao procurar informações. O ChatGPT também está ganhando popularidade como ferramenta de pesquisa, com mais de 10% das pessoas usando-o para consultar informações como conselhos sobre finanças pessoais. [...] O TikTok se tornou popular como mecanismo de busca porque entrega conteúdo de forma única e personalizada – por meio de vídeos curtos e informativos que contam pequenas histórias. Uma grande parte do apelo do TikTok é como ele adapta o conteúdo aos interesses de cada usuário (Southern, 2024, tradução nossa).

Keinonen (2008) evidenciou sobre o movimento na indústria ao expor o contraste com as tendências anteriores aos avanços tecnológicos – em meados dos anos 2000 - que emergiu da necessidade de adaptar tecnologias profissionais complexas para uso prático pelo público em geral. Essa mudança levou à aplicação da ciência cognitiva e da psicologia para entender e melhorar a interação humano-computador em ambientes realistas. O movimento também se concentrou na simplificação do processo de design, incorporando práticas racionalizadas.

Esse campo evoluiu ainda mais nas últimas décadas, expandindo-se além do escopo da otimização de design humano-computador. Atualmente, abrange-se uma gama mais ampla de interesses, buscando alinhar novas tecnologias interativas com comunicação entre humanos e requisitos organizacionais, bem como perspectivas sociais e emocionais. Essa mudança destaca a crescente importância de considerar vários aspectos da interação e experiência humana ao desenvolver e implementar tecnologias interativas.

Arrisco ao afirmar que o foco das abordagens tradicionais de consumo, não só no digital, mudou. Meira (2019) já apontava em uma época pré-pandemia uma transição da mídia em massa e propagandas genéricas para uma abordagem mais centrada na singularidade de necessidades e preferências dos usuários.

Keinonen (2008) afirma que os produtos e serviços eram desenvolvidos com base em premissas e tendências gerais do mercado. A evolução das tecnologias digitais e a proliferação de dados permitiram as organizações coletarem e analisarem o *feedback* e o comportamento do usuário, em seguida, voltar para a análise dos resultados obtido e tomar decisões mais direcionadas ao comportamento do usuário.

O *feedback* trouxe o foco no usuário como uma forte influência para a evolução dos serviços. Esse enfoque também abriu caminho para a cocriação e a inovação colaborativa, pois as organizações se envolvem ativamente em projetar produtos e serviços para analisar a motivação e o comportamento desse usuário, de modo a fornecer recomendações personalizadas e melhorar a experiência na utilização do produto/serviço.

Para um funcionamento adequado à prontidão das respostas, esses serviços personalizados utilizam bases robustas de Inteligência Artificial e *Big Data*. Para Ribeiro (2014), *Big Data* engloba a vasta quantidade de informações que nos rodeia no dia a dia, abrangendo a diversidade informacional, a velocidade com que são gerados e transmitidos, bem como a qualidade dos dados e das informações envolvidas. A Inteligência Artificial auxilia no desenvolvimento de ferramentas que facilitam o treinamento efetivo dos termos na organização informacional voltada ao viés da singularidade e das preferências de cada público.

A integração de *Big Data* e IA nesses serviços também permite aprendizado e melhoria contínua. À medida que mais dados são coletados e analisados, os algoritmos se tornam mais refinados, levando a recomendações mais precisas ao longo do tempo. Além disso, esses serviços podem se adaptar às mudanças nas

preferências e comportamentos do usuário, garantindo que as recomendações permaneçam relevantes e atualizadas.

Sharma e Lijuan (2015) realizaram um estudo para investigar essa relação e descobriram que a qualidade do serviço *online* influencia significativamente a satisfação do usuário. Do ponto de vista dos clientes, suas expectativas precisam ser atendidas e suas interações com os produtos devem ser uma experiência positiva, para que as devidas repercussões sejam consideradas um sucesso.

O sucesso não é determinado apenas pela funcionalidade ou design, mas também é significativamente influenciado pela resposta emocional que evoca nos clientes. Essa perspectiva entra em acordo com a visão apresentada por Don Norman (2013), que expõe que uma experiência positiva do usuário deixa uma impressão duradoura e promove confiança e lealdade.

Dessa forma, ao levar em conta o público-alvo e ajustar o conteúdo e recursos correspondentes, estabelecer uma reputação de marca sólida e, por fim, atingir a meta de promover uma experiência de usuário perfeita, os usuários experimentam a impressão de serem ouvidos, compreendidos e satisfeitos formando um elo sólido de confiança.

4.1 A experiência do usuário

Embora a experiência do usuário seja um tema em alta no mundo digital, vale destacar que estudos com foco no usuário já existiam bem antes, como um campo interdisciplinar explorado por diversas áreas, incluindo a Biblioteconomia e Ciência da Informação.

Baptista e Cunha (2007) afirmam que há mais de 40 anos os estudos de usuários consolidam seus objetivos: aprimorar produtos e serviços informacionais e mapear o fluxo da informação. As autoras também relatam sobre os estudos de abordagem qualitativas para estudar sobre o usuário, dentre eles: a usabilidade.

O termo usabilidade é frequentemente usado por pesquisadores em experiência do usuário e arquitetura da informação. Embora as autoras mencionem uma possível substituição na nomenclatura, acredito que seja mais uma questão de tradução que reforça o significado em português, sem substituí-lo. O objetivo da usabilidade é tornar a interação humano-computador amigável e fácil para o usuário.

Os estudos de usabilidade têm crescido na procura de soluções para sistemas de informação automatizados, com as páginas da Web e outros tipos de interação via sistemas amigáveis. O termo usabilidade substituiu o termo *user friendly* que tinha conotações subjetivas. A usabilidade pode ser mensurada de acordo com os atributos ergonômicos do produto, em termos do esforço mental e atitudes dos usuários e pela forma como os usuários interagem com o produto e sua aceitação (Bevan; Macleod, 1994 *apud* Baptista; Cunha, 2007).

A usabilidade transcende o mundo digital, encontrando aplicação em espaços físicos como, por exemplo, as bibliotecas. A disposição das estantes, a distância entre elas, a iluminação, a ventilação e a organização de todo o espaço não visam facilitar somente o trabalho do bibliotecário, mas proporcionar ao usuário uma experiência amigável, facilitando a busca e o encontro de informações.

Outro exemplo é encontrado nos supermercados. Desde a disposição das estantes até a altura em que os produtos são organizados, cada detalhe pode influenciar a experiência do consumidor e sua decisão de compra. A forma como diferentes seções do supermercado é organizada, como os itens para bolos próximos aos itens de festa, facilita a experiência de compra.

No ambiente de supermercado, a disposição estratégica dos produtos cria uma experiência de compra fácil e até mesmo influenciável. Por exemplo, a fila do caixa frequentemente possui uma estante com itens tentadores a preços acessíveis. Isso demonstra como o estudo de experiência do usuário vai além da internet, impactando a vida cotidiana das pessoas.

Essa facilidade de uso serve de inspiração para profissionais que buscam melhorar a experiência em plataformas digitais, visando uma experiência de uso mais positiva. Para compreender o significado e a construção da experiência do usuário (UX), é essencial explorar suas origens, propósito e estado atual.

As máquinas não são pessoas. Elas não podem se comunicar e entender da mesma forma que nós. Isso significa que os seus designers têm uma obrigação especial de garantir que o comportamento das máquinas seja compreensível para as pessoas que interagem com elas. Uma verdadeira colaboração requer que cada parte faça um esforço para se adaptar e entender a outra. Quando colaboramos com máquinas, são as pessoas que precisam fazer todas as adaptações. Por que a máquina não poderia ser mais amigável? A máquina deveria aceitar o comportamento humano normal, mas assim como as pessoas frequentemente avaliam subconscientemente a precisão do que está sendo dito, as máquinas deveriam julgar a qualidade da informação que lhes é fornecida, neste caso para ajudar seus operadores a evitar erros graves devido a simples deslizos. Hoje em dia, insistimos que as pessoas ajam de forma anormal, se adaptando às exigências peculiares das máquinas, o que inclui sempre fornecer informações precisas e corretas. Os humanos são particularmente ruins nisso, ainda assim, quando não

conseguem atender às exigências arbitrárias e desumanas das máquinas, chamamos isso de erro humano. Não, é um erro de design (Norman, 2013, p. 67, tradução nossa).

O conceito de UX surgiu como resposta à crescente complexidade no uso da tecnologia e a necessidade de aprender a manusear interfaces complexas, que possuem o potencial em ser mais intuitivas e amigáveis. Exercendo o papel de prover essa facilidade para os usuários, a experiência do usuário tornou-se um aspecto crucial do processo do design das plataformas, e aqui o design vai além da interface, como o desenvolvimento de todo o projeto.

Um dos pioneiros nesse campo é Don Norman, um renomado cientista cognitivo, designer e escritor, que fez contribuições significativas para a compreensão e melhoria da experiência do usuário. No livro "*The Design of Everyday Things*", em sua versão atualizada e revisada, de 2013, Norman enfatiza a importância de projetar para humanos e defende uma abordagem centrada no usuário. Remete que um bom design não deve ser apenas reduzido à elaboração de uma estética agradável, mas, de fato, sobre melhorar e expandir a experiência geral de uso dos objetos e informações, considerando os estímulos psicológicos e cognitivos positivos do comportamento humano (Norman, 2013).

Rocha, Paula e Duarte (2016) destacam que os processos cognitivos do estudo de usuário são distribuídos no tempo e são moldados pelas experiências passadas dos indivíduos, permitindo que interpretem o presente e antecipem o futuro com base em lembranças de situações similares. Os autores referenciam Hutchins (2000), que ressalta o papel dos artefatos como mediadores na interação entre os indivíduos e o ambiente.

Seu estudo aponta que, no mundo contemporâneo, a proximidade física não é mais crucial devido às tecnologias de informação e comunicação, que facilitam a distribuição social do trabalho cognitivo. Além disso, salientam que o estudo dos processos cognitivos distribuídos está intimamente ligado ao estudo da cultura, pois as experiências culturais passadas moldam a interpretação do presente e a percepção do futuro.

Bons designers nunca começam tentando resolver o problema que lhes é dado: eles começam tentando entender quais são os verdadeiros problemas. Como resultado, em vez de convergirem para uma solução, divergem, estudando as pessoas e o que elas estão tentando realizar, gerando ideia após ideia. (...) Os designers resistem à tentação de saltar imediatamente

para uma solução para o problema declarado. Em vez disso, primeiro gastam tempo determinando qual questão básica e fundamental (raiz) precisa ser abordada. Eles não tentam procurar uma solução até terem determinado o problema real e, mesmo assim, em vez de resolver esse problema, param para considerar uma ampla gama de soluções potenciais. Só então eles finalmente convergirão para a sua proposta (Norman, 2013, p. 218 a 219, tradução nossa).

Também utilizando a abordagem voltada ao cognitivo, Norman introduziu os conceitos de *signifiers* e *affordances*, e como eles influenciam e impactam no uso e percepção dos objetos.

Affordances se refere aos aspectos e sinais feitos dentro de um projeto de design para facilitar a interação e fornecer dicas sobre como usar ou manipular um objeto ou ambiente no qual ele se encontra. As propriedades do *affordance* podem ser físicas, perceptivas ou até mesmo conceituais. Seu principal objetivo é orientar os usuários para o uso correto da funcionalidade pretendida.

Os *signifiers* são elementos que orientam os usuários sobre como interagir com um determinado objeto ou sistema. Esses elementos são projetados para serem facilmente perceptíveis e intuitivos, permitindo que os usuários entendam rapidamente como interagir com a interface.

Ainda sobre os *signifiers*, podem assumir várias formas, como ícones, botões, rótulos ou sons, dependendo do contexto e do meio utilizado. A ausência ou ambiguidade de *signifiers* pode levar a problemas de usabilidade, pois os usuários podem se esforçar para entender como interagir com um objeto ou sistema.

Portanto, considerar diversas dimensões, como a cultural, social, histórica, emocional e subjetiva, é fundamental para compreender como os indivíduos atribuem significado à informação e constroem conhecimento. Deste modo, é importante analisar a raiz das questões além do problema declarado para considerar as diferentes dimensões de cada grupo de pessoas.

À medida que a tecnologia avançava e se tornava mais integrada em nosso dia a dia, os usuários começaram a esperar experiências agradáveis em várias plataformas e dispositivos. Portanto, o foco de UX passa a ser nas necessidades, desejos e limitações dos usuários, oriundos das interações com o próprio sistema, e que sejam intuitivas, de modo a proporcionar uma experiência positiva.

Uma estratégia eficiente de experiência do usuário consiste em compreender os comportamentos dos usuários, conduzir pesquisas com eles, bem como testar e aprimorar os designs de forma interativa, de modo a satisfazer suas expectativas. O

propósito é eliminar os problemas enfrentados, como a frustração, e simplificar os processos, visando melhorar de maneira geral a satisfação dos clientes.

A pesquisa inicial para entender a natureza do próprio problema faz parte da disciplina da pesquisa em design. Note que esta é uma pesquisa que usará os produtos em questão. Não é o tipo de pesquisa que os cientistas realizam em seus laboratórios, tentando encontrar novas leis da natureza. O pesquisador de design irá até os potenciais clientes, observando suas atividades, tentando compreender seus interesses, motivações e necessidades reais. A definição do problema para o design do produto surgirá a partir desse profundo entendimento dos objetivos que as pessoas estão tentando alcançar e das barreiras que possam encontrar em seu ambiente natural, em suas vidas normais, onde o produto ou serviço a ser projetado será realmente utilizado. (...) é essencial entender as situações reais que eles enfrentam, e não alguma experiência isolada e pura (Norman, 2013, p. 222, tradução nossa.)

Quando os usuários têm uma experiência positiva com um produto ou serviço, é mais provável que se tornem consumidores constantes. Além disso, uma experiência bem projetada pode reduzir os custos de suporte ao cliente, pois os usuários podem navegar facilmente e encontrar soluções por conta própria, também estimulando a autonomia e acelerando a curva de aprendizado de uso. Empresas que priorizam UX investem no entendimento de seu público-alvo, mantêm-se atualizadas com as últimas tendências e tecnologias de design e em uma melhoria contínua de seus produtos e serviços.

Com a participação ativa dos usuários e a centralização de suas necessidades, é necessário organizar o que pode ser percebido como um caos, decorrente da natureza contínua da experiência e da complexidade da personalização. A taxonomia surge como uma ferramenta fundamental para auxiliar na organização e disponibilização dos produtos, atendendo a essas necessidades. Em teoria, a taxonomia é de fácil construção, acessível e potencializa diversos sistemas, como os de recomendação, temas que serão abordados nos próximos capítulos.

5 O VALOR DA TAXONOMIA

Uma taxonomia é uma ferramenta usada para classificar conceitos em uma ordem hierárquica, representando um conhecimento específico. Isso significa que ela categoriza informações com base em um contexto delimitado, atribuindo-lhes significado. Funciona como uma classificação para diversos propósitos, refletindo um ponto de vista particular.

Surgiu inicialmente através da área da biologia para a classificação dos seres vivos e não vivos. Mendes e Bentes Pinto (2019) observam que o método classificatório de acordo com as semelhanças das características do objeto na biologia está intimamente ligado às linguagens documentárias, e, por conseguinte, à Biblioteconomia e à Ciência da Informação. Relatam que, ao nomear os elementos, não apenas os compreendemos em suas respectivas semelhanças, mas isso facilita a organização do conhecimento registrado, assim como o acesso e a recuperação da informação.

Para Bailey (1994), as tipologias e taxonomias são métodos de classificação comumente usados em ciências naturais, sociais e humanas. Tipologias são usadas para categorizar objetos ou eventos em grupos discretos com base em suas características, enquanto taxonomias são usadas para classificar objetos ou eventos em estruturas hierárquicas com base em seus relacionamentos.

Para ilustrar a hierarquia da classificação, podemos usar a Biologia e a metáfora de uma árvore, onde as grandes raízes representam a base da taxonomia com o contexto e a governança, sustentando toda a estrutura. Assim como uma árvore, a taxonomia começa com classes generalistas que se ramificam em subclasses.

As classes generalistas, equivalente ao caule da árvore, são amplas o suficiente para abranger todas as outras categorias, mas não precisam ser extensas. Servem como base para as demais informações e devem ser claras e assertivas o bastante para detalhar as subclasses.

Por isso, é crucial que contenha poucas classes e conceitos gerais em vez de extensas categorias específicas, que podem gerar ambiguidades e problemas de conceitualização. É essencial que as classes sejam claras em seu significado para evitar dúvidas e garantir uma categorização eficiente.

Embora esses conceitos sejam inerentes e até intuitivos para a área da Ciência da Informação, é comum encontrar taxonomias produzidas por outras áreas correlatas que não seguem essa linha de raciocínio. Isso resulta em uma desmedida base de classes, misturando categorias gerais com específicas, criando uma complexidade desnecessária no sistema classificatório.

Essa abordagem desarticulada pode causar uma série de problemas, tornando a estrutura confusa e difícil de corrigir. Ajustes em um ponto específico podem causar rupturas em outros, criando uma espécie de "bola de neve" na tentativa de reparar a taxonomia.

Por ser acessível e adaptável para diferentes contextos informacionais, a taxonomia pode ser aplicada para diferentes propósitos. Aqui estão alguns exemplos:

Treinamento de Inteligência Artificial: pode ser empregada em sistemas de Aprendizado de Máquina (ML), utilizando Processamento de Linguagem Natural (PLN).

Na prática, Jones (2019) aponta o impacto no processo de treinamento de IA usando o Processamento de Linguagem Natural do Google para analisar o conteúdo, que divide em entidades. Essas entidades são destacadas em cores, cada uma com um número abaixo. Esses números representam classificações relativas com base na importância da entidade dentro do artigo.

Apesar de não mencionar explicitamente a taxonomia, fica visível o uso da ferramenta para a classificação, principalmente porque a empresa Google produz e vende taxonomias para o mercado.

Auxílio em sistemas de recomendação: facilita o funcionamento dos sistemas de recomendação, no entendimento e na interpretação dos *clusters*, pois fornece uma estrutura navegável. Viabiliza a recuperação eficiente de informações, permitindo que os usuários pesquisem e acessem grupos ou categorias de dados específicos, servindo como guia que auxilia os usuários a explorarem e analisarem os dados de forma mais sistemática e estruturada.

A importância da taxonomia destaca-se ainda por sua capacidade de organizar e categorizar pontos de dados com base em seus relacionamentos e semelhanças. Ao agrupar pontos de dados relacionados, a taxonomia permite a criação de *clusters* mais significativos, que capturem padrões e estruturas subjacentes nos dados (Pinterest Engineering, 2020).

Segmentar os usuários, de forma eficiente e precisa, em grupos coesos é fundamental para o aumento da receita no setor de tecnologia e vendas online (Ramesh *et al.*, 2013). Assim, a taxonomia auxilia na redução da complexidade do agrupamento de itens semelhantes, colaborando com a análise e interpretação dos dados e informações.

Este ponto será abordado no capítulo seguinte, com ainda mais detalhes.

Marketing e Publicidade: permite a criação de catálogos de produtos estruturados e de autoatendimento, campanhas segmentadas e personalização de vendas e anúncios. Ao construir audiências de forma precisa, o setor economiza recursos financeiros e alcança o público-alvo adequado para suas campanhas.

Na prática: Freni (2024) revela em seu artigo um caso em que um executivo de marketing compartilhou que até 65% dos dados de sua equipe são imprecisos devido à má implementação da taxonomia e à falta de padrões de dados. Apontando assim a importância de uma boa taxonomia e seu respectivo impacto no negócio.

Vimos contextos em que a taxonomia pode ser uma ferramenta potencializadora, e algumas empresas a utilizam amplamente, como o Pinterest, Meta e Google, para criar campanhas "self-service". O New York Times possui um sistema robusto para marketing e publicidade. Além disso, a Amazon a utiliza para melhorar seus sistemas de recomendação de produtos, entre outras aplicações.

Contudo, essa ferramenta não é utilizada da mesma maneira por todos. Reforço novamente que, apesar de ser um senso comum dentro da área da ciência da informação, apontar a necessidade de diferenciar as taxonomias e identificar quais contextos elas melhor se encaixam é essencial para áreas correlatas.

Um exemplo para ilustrar é que se torna impossível realizar todo um processo de aprendizagem de máquina usando só SQL. Portanto, não é viável usar uma única taxonomia com um único propósito para todos os contextos possíveis. A seguir, veremos as características de três tipos principais de taxonomia no contexto deste trabalho: a tradicional (não relacionada à biologia, mas à área da Ciência da Informação), a de Audiência e a de PLN.

5.1 Comparação entre Taxonomia tradicional e Taxonomia de audiência

Nesta seção, serão exploradas as diferenças entre as taxonomias tradicionais e taxonomias de audiência. Embora ambas sirvam ao propósito de classificação,

divergem em abordagem e foco. Uma taxonomia tradicional é comumente usada em campos como Biologia e Biblioteconomia.

Conforme definido por Klein (2017), o objetivo de uma taxonomia tradicional é criar um sistema de classificação sistemático e abrangente, que cubra todas as entidades possíveis dentro de um determinado domínio. Tem como característica principal fornecer uma estrutura padronizada que pode ser aplicada e compreendida universalmente.

Por outro lado, uma taxonomia de audiência se concentra na categorização de informações com base em preferências, interesses ou necessidades individuais. Reconhece-se, portanto, que as pessoas têm perspectivas diferentes e seus interesses podem variar. Na taxonomia de audiência, não se considera a universalidade, mas as necessidades do usuário e seus interesses particulares.

Uma taxonomia de audiência permite que os usuários personalizem seu sistema de classificação, selecionando categorias que se alinhem com suas vontades específicas. Desse modo, fornece-se uma abordagem mais flexível e dinâmica para organizar as informações, permitindo que os usuários naveguem e acessem conteúdos relevantes e significativos para eles.

Segundo Aldayri e Albattah (2022), a taxonomia de audiência é um aspecto crucial das estratégias de marketing e comunicação, pois auxilia a identificar e entender melhor o público-alvo, possibilitando a criação de conteúdos mais personalizados e relevantes.

Ao analisar essa taxonomia, os profissionais de marketing obtêm informações sobre as diferentes necessidades, motivações e preferências de vários segmentos de público, permitindo que desenvolvam campanhas direcionadas que ressoam com grupos específicos.

Essa abordagem aumenta a eficácia dos esforços e ajuda as organizações a construir conexões mais fortes com seu público, transmitindo mensagens adaptadas às suas características e interesses exclusivos.

Desse modo, a principal diferença entre uma taxonomia tradicional e uma taxonomia de audiência está em seu propósito e design. A taxonomia tradicional visa criar um sistema abrangente e objetivo, que pode ser aplicado uniformemente em diferentes contextos. Além de centralizar na criação de uma estrutura padronizada e universal, que facilite a recuperação de informações e garanta a consistência na categorização. Por outro lado, a taxonomia de audiência concentra o foco na

categorização centrada no usuário, permitindo que os indivíduos encontrem e explorem informações com base em suas preferências e interesses pessoais.

Outra distinção significativa é o nível de flexibilidade e adaptabilidade. As taxonomias tradicionais são tipicamente estáticas e requerem atualizações periódicas para incorporar novas entidades ou mudanças no conhecimento. Portanto, são projetados para serem estáveis e consistentes ao longo do tempo.

Em contraste a essas, as taxonomias de audiência são mais dinâmicas, uma vez que respondem às necessidades individuais. Os usuários podem adicionar ou modificar categorias com base em seus interesses em evolução, garantindo que a taxonomia permaneça relevante e personalizada.

Um exemplo de taxonomia de audiência é disponibilizado pelo Pinterest (Pinterest Engineering, 2020). A empresa utiliza um sistema de taxonomia para organizar e selecionar tópicos e interesses populares com eficiência para fins de publicidade direcionada.

Conforme Pinterest Engineering (2020), entender as tendências atuais requer uma compreensão mais profunda do conteúdo que os *Pinner*s estão procurando, bem como das categorias às quais os *Pins* estão associados.

Para conseguir isso, o *Pinterest* desenvolveu um sistema de gerenciamento de conhecimento, também conhecido como *knowledge graph*, baseado em taxonomia.

Esse sistema permite a compreensão eficiente do conteúdo, aproveitando a estrutura da taxonomia. Ao categorizar e organizar a vasta quantidade de conteúdo na plataforma, o *Pinterest* reúne informações valiosas sobre o comportamento do usuário, tendências emergentes e evolução dos gostos do consumidor.

O *status* único da plataforma - que é usada por mais de 320 milhões de usuários em todo o mundo para salvar ideias e fazer planejamentos diversos - fornece à empresa uma riqueza de dados para análise.

De tarefas diárias - como decidir o que vestir para trabalhar ou comidas para cozinhar - a eventos significativos da vida - como decorar uma nova casa ou planejar uma viagem - os usuários salvam mais de 200 bilhões de ideias em mais de 4 bilhões de painéis.

Essa enorme coleção de ideias salvas serve como uma rica fonte de informações, oferecendo ideias sobre tendências emergentes e sinais iniciais do comportamento do consumidor. Ao alavancar o sistema de gerenciamento de

conhecimento baseado em taxonomia e analisar a grande quantidade de ideias salvas, o *Pinterest* pode ficar na frente da compreensão do que seus usuários estão procurando e como suas preferências evoluem.

Esse entendimento dos interesses dos usuários permite que a plataforma ofereça oportunidades de publicidade direcionada aos anunciantes, garantindo que os usuários recebam conteúdo relevante e que se alinhem com os interesses e necessidades.

No geral, o sistema de taxonomia de audiência e a abordagem de gerenciamento de conhecimento permitem que a plataforma ofereça uma experiência personalizada e envolvente para seus usuários. Ao mesmo tempo, fornece informações valiosas sobre o comportamento do consumidor e as tendências emergentes.

O valor de uma taxonomia de audiência é o potencial para fornecer informações valiosas sobre os interesses de indivíduos, grupos ou organizações. Essa subjetividade permite que uma taxonomia de audiência seja mais flexível e adaptável, atendendo às diversas necessidades e gostos de diferentes indivíduos.

Ambas as taxonomias servem a propósitos importantes em diferentes domínios. A taxonomia tradicional ajuda cientistas e pesquisadores a entender o mundo natural e sua biodiversidade; uma taxonomia de interesse aprimora a experiência do usuário ao fornecer categorização e recomendações personalizadas.

5.2 Comparação entre taxonomia tradicional e taxonomia PLN

O processamento da linguagem natural (PLN) é uma área da computação que aborda a comunicação humana. Seu objetivo é permitir que os computadores compreendam e se comuniquem em linguagem humana, abrangendo diferentes níveis de entendimento e geração de linguagem, incluindo sons, palavras, frases e discursos.

Esses níveis, conforme identificados por Marco Gonzalez (2003), incluem: Fonético e Fonológico: relacionado à forma como as palavras se relacionam com seus sons; Morfológico: diz respeito à construção das palavras a partir de unidades de significado primitivas; Sintático: trata das relações entre as palavras em uma sentença; Semântico: envolve o relacionamento das palavras com seus significados; Pragmático: lida com o uso de frases e sentenças em diferentes contextos.

De acordo com Marco Gonzalez (2003), métodos estatísticos desempenham um papel significativo no PLN, ajudando em tarefas como resolução de ambiguidade, etiquetagem e aquisição de conhecimento lexical. Enfatizam que esses métodos são especialmente valiosos quando vistos à luz da teoria da informação, pois a probabilidade de termos ocorrerem juntos em relação às suas ocorrências isoladas fornece informações cruciais. Essa medida é útil, por exemplo, para calcular a similaridade entre palavras, algo fundamental para o funcionamento eficaz do PLN.

Apesar das pesquisas dos autores se concentrarem principalmente na ontologia, fazem uma menção à taxonomia e sua utilidade, o que se alinha com a abordagem deste projeto. Em uma taxonomia, há regras de formação que determinam os relacionamentos válidos entre informações, baseados na ideia de hierarquia de objetos, em que características são herdadas de ancestrais para descendentes. As palavras presentes nos léxicos são usadas para fazer referência aos conceitos apresentados pelos autores em uma ontologia (Marco Gonzalez, 2003). Nesse sentido, a hierarquização da taxonomia também desempenha um papel importante na organização da informação, algo que pode ser aplicado de forma relevante ao PLN.

O conceito de taxonomia tradicional visa estabelecer um sistema de classificação sistemático e abrangente, dentro de um domínio específico, fornecendo uma estrutura padronizada aplicável e compreendida universalmente.

Cambria e White (2014) introduzem o conceito de PNL taxonômica, que se concentra na construção de taxonomias universais para compreender a semântica hierárquica associada a expressões de linguagem natural. Essas taxonomias abrangem conceitos, instâncias, atributos, valores e relacionamentos, com ênfase particular em representações de conhecimento subjuntivas baseadas em relacionamentos 'É uma', 'Pertence a'.

Embora a taxonomia PLN ofereça uma perspectiva única na representação contextual, observa-se que outras abordagens, como redes bayesianas, ou seja, redes de representação que exploram de forma explícita o papel do conhecimento prévio, combinando evidências sobre a probabilidade de eventos ocorrerem, e as redes semânticas, que têm a responsabilidade de desempenhar diversas funções, como lembrar, comparar, generalizar, exemplificar, fazer analogias, simplificar, prever, e entre outras (Cambria; White, 2014), também são utilizadas na pesquisa de PNL.

À medida que o campo da PNL avança, a exploração e comparação dessas diferentes abordagens contribuem para o progresso da representação e compreensão

do conhecimento no processamento de linguagem natural. Essa abordagem visa melhorar a compreensão do texto, organizando as informações hierarquicamente, permitindo a recuperação e inferência de conhecimento eficiente.

Segundo Cambria e White (2014), a PNL taxonômica utiliza taxonomias para representar e organizar o conhecimento de forma semelhante à forma como os humanos organizam as informações em suas mentes.

Ao empregar taxonomias, a PNL taxonômica fornece um mecanismo para codificar relacionamentos hierárquicos entre conceitos, facilitando a identificação de semelhanças e diferenças entre eles. Essa abordagem é particularmente útil em várias tarefas de PLN, como recuperação de informações, resposta a perguntas e análise de sentimentos, pois permite que o sistema tome decisões mais equilibradas e sutis sobre o contexto.

Além disso, a PNL taxonômica aumenta a capacidade de interpretação dos sistemas de PNL, uma vez que as estruturas hierárquicas permitem representações de conhecimento mais transparentes e explicáveis.

Em contraste das taxonomias tradicionais, a taxonomia PLN centraliza na extração e utilização do conhecimento taxonômico para aprimorar a compreensão da linguagem e a recuperação de informações, assim como também visam organizar e categorizar várias entidades, com base em suas características compartilhadas e relacionamentos.

A taxonomia PLN pode melhorar a precisão dos sistemas de PLN, incorporando informações taxonômicas no processo de compreensão da linguagem. Ao alavancar o conhecimento taxonômico, como relacionamentos hierárquicos entre conceitos, a taxonomia PLN potencializa o desempenho de várias tarefas, incluindo classificação de texto, análise de sentimento e recuperação de informações.

Ademais, a taxonomia PLN oferece uma vantagem única ao alavancar o conhecimento taxonômico, que pode fornecer uma compreensão mais abrangente da linguagem e facilitar o processamento de linguagem mais preciso.

A taxonomia tradicional e a taxonomia PLN oferecem diferentes perspectivas sobre os sistemas de classificação no campo do processamento de linguagem natural. No geral, o estudo da taxonomia PLN fornece informações valiosas sobre como os sistemas de classificação podem ser adaptados especificamente para a análise de dados de linguagem natural, contribuindo para os avanços no campo da linguagem computacional.

Ao focar nas relações hierárquicas entre conceitos e suas subcategorias, a taxonomia PLN apresenta um caminho valioso para explorar o conhecimento de integração e sua aplicação em sistemas de agrupamento, de modo a permitir o aumento do desenvolvimento de técnicas de clusterização mais eficientes.

Portanto, essa abordagem melhora a precisão e a eficácia da recuperação de informações, bem como potencializa a compreensão geral e contextual dos dados textuais.

À medida que o campo da taxonomia PLN evolui, amplia-se o potencial para o avanço de vários domínios, incluindo recuperação de informações, análise semântica e sistemas de recomendação.

Quadro 1 - Características gerais das taxonomias: tradicional, de audiência e PLN

Tradicional	Audiência	PLN
Sistema de classificação sistemático e abrangente	Categorização de informações com base em preferências, interesses ou necessidades individuais.	Mecanismo para codificar relacionamentos hierárquicos entre conceitos
Aplicabilidade uniforme em diferentes contextos	Permite personalização por parte do usuário	Permite representações de conhecimento mais transparentes e explicáveis.
Estáticas e atualizações periódicas	Auxilia a identificar e entender melhor o público-alvo	Aprimora a compreensão da linguagem e a recuperação de informações
	Dinâmicas	Organizar e categorizar várias entidades

Fonte: Elaborado pela autora (2024).

5.3 Prós e contras de uma taxonomia para um sistema recomendação

Na perspectiva da análise de qualidade, examinar todas as combinações de grupos possíveis, em um grande volume de dados, pode ser uma tarefa árdua e demorada. Para superar esse desafio, pesquisadores e cientistas de dados desenvolveram diversas técnicas que podem auxiliar na formação de grupos citados

por Doni (2004) em seu trabalho. Uma dessas técnicas citadas por ele é a análise de agrupamento proposta por Zaiane (2003).

Não obstante, as características são: a alta dimensionalidade; ser “escalável”; lidar com diferentes tipos de dados; definir agrupamentos de diferentes tamanhos e formas; mínimo de conhecimento para determinação dos parâmetros de entrada; robusto à presença de ruído; e apresentar resultado consistente independente da ordem em que os dados são apresentados.

Métodos taxonômicos são métodos empíricos. Eles geralmente começam com um conjunto de dados de objetos empíricos (como indivíduos) medidos em várias variáveis. Várias técnicas, geralmente quantitativas, são usadas para agrupar os casos por semelhança geral nas variáveis (Bailey, 1994, tradução nossa).

Realizar uma análise de *cluster* cuidadosa requer a utilização de métodos que podem ir além das características apontadas na pesquisa dos autores. Neste trabalho, foi-se adicionado às características particulares da taxonomia inata das ciências sociais que acredito que possam auxiliar no processo de análise de cluster.

Essas características inerentes podem incluir:

1. **Organização de dados otimizada:** capacidade de localizar e acessar facilmente dados específicos. Contribui para a validação e confiabilidade das descobertas através dos dados, permitindo conclusões significativas através de decisões consolidadas.
2. **Aprimoramento da pesquisa:** ao estruturar os dados em categorias distintas, com base em critérios relevantes, a taxonomia permite uma recuperação mais refinada e eficiente das informações desejadas.
3. **Análise de dados otimizada:** por meio da organização sistemática de dados em categorias distintas, uma taxonomia possui a capacidade de otimizar a análise de dados.
4. **Visualização de dados aperfeiçoada:** por meio do arranjo sistemático de dados em categorias específicas, uma taxonomia facilita a organização e estruturação de informações, bem como contribui para a eficácia geral da visualização de dados. Isso pode ser especialmente útil ao lidar com grandes quantidades de dados.

Por outro lado, as desvantagens de usar uma taxonomia para um sistema de agrupamento incluem:

1. **Super Categorização:** uma taxonomia pode levar à supercategorização, que pode ocasionar uma análise de dados imprecisa devido aos possíveis fatores - como ambiguidades.
2. **Imprecisão:** a não definição precisa das categorias pode acarretar análises enviesadas.
3. **Dificuldade na mudança de categorias:** a modificação de uma taxonomia pode apresentar desafios se as categorias não tiverem sido definidas com exatidão, principalmente quando surgem problemas de supercategorização e imprecisão. Nesses casos, qualquer tentativa de alteração da taxonomia pode agravar os problemas existentes e, conseqüentemente, ampliar negativamente a sua visibilidade. Portanto, a taxonomia perde seu papel de estrutura e passa a ser associada a uma infinidade de questões específicas, em vez de servir como uma entidade discreta.

Para que esse modelo funcione bem, é crucial ter um sistema sólido de análise de dados, que se baseie nas informações fornecidas pelo usuário. Esse sistema envolve organizar os dados em grupos e categorias relevantes que facilitem a oferta de recomendações úteis para o usuário.

Nesta seção, foi explorado os temas relacionados a cada um desses tópicos, o que proporcionou uma compreensão abrangente do cenário atual, incluindo seus aspectos positivos e negativos.

Traçamos uma linha do tempo desde o surgimento da web 2.0 e suas conseqüências sociais, destacando o papel central dos usuários na criação de conteúdo; até chegarmos à análise do cenário atual, no qual o comportamento dos usuários influencia as decisões e orienta a geração de informações que impactam produtos, instituições e empresas.

Com esse conhecimento, estamos prontos para examinar e questionar: é viável considerar a taxonomia uma ferramenta importante para o processo de recomendação? A resposta é: sim, pois a taxonomia desempenha um papel crucial no processo de clusterização de dados.

Justifica-se essa afirmação ao considerarmos que a configuração da classificação, onde os itens são agrupados em categorias mais amplas e depois divididos em subcategorias, com base em características compartilhadas, sendo essa ordenação fundamental no processo de clusterização, é justamente um benefício que se obtém com o uso de taxonomias.

6 CLUSTERIZAÇÃO E RECOMENDAÇÃO DE CONTEÚDO

A organização da informação é o processo de estruturar e organizar os dados de forma a torná-los mais fáceis de encontrar, acessar e usar. É uma parte importante do gerenciamento de informações e é essencial para qualquer recurso informacional que precise armazenar e recuperar informações de forma rápida e eficiente.

É possível, ao organizar as informações, aproveitar ao máximo os recursos dispostos e garantir que os dados estejam seguros, atualizados e precisos. Dessa forma, torna-se mais fácil de encontrar e usar os dados; economiza tempo e dinheiro; facilita a localização e o uso das informações; ajuda a garantir que as informações sejam precisas e atualizadas; e favorece a garantia que os dados estejam seguros e protegidos contra acesso não autorizado.

Para esse processo funcionar, deve-se envolver a criação de uma estrutura lógica, consistente e fácil de entender. Essa estrutura pode ser baseada em categorias, tópicos ou outros critérios que facilitem a localização da informação necessária. Por meio de um sistema de agrupamento, conhecido como clusterização - termo este que será usado no decorrer deste trabalho para se referir ao processo de agrupamento -, que seja capaz de agrupar informações relacionadas. Isso será discutido adiante neste texto

Essas estruturas lógicas e consistentes podem ser hierárquicas, ou seja, as informações são organizadas em uma série de níveis, cada nível contendo informações mais específicas do que o anterior.

Desse modo, a opção de estrutura hierárquica que pode agrupar informações em categorias é conhecida como taxonomia (Silveira *et al.*, 2021). Nessas estruturas, as informações são divididas em categorias e subcategorias – cada categoria contém informações mais específicas do que as anteriores. Não obstante, permite-se uma busca mais eficiente, pois as informações relacionadas podem ser localizadas rapidamente por meio de buscas nas categorias e subcategorias.

A taxonomia expandiu para além do campo da biologia, limitando-se a classificação de seres vivos e não vivos, e foi adaptada para classificar os dados e as informações disponíveis que temos hoje em dia.

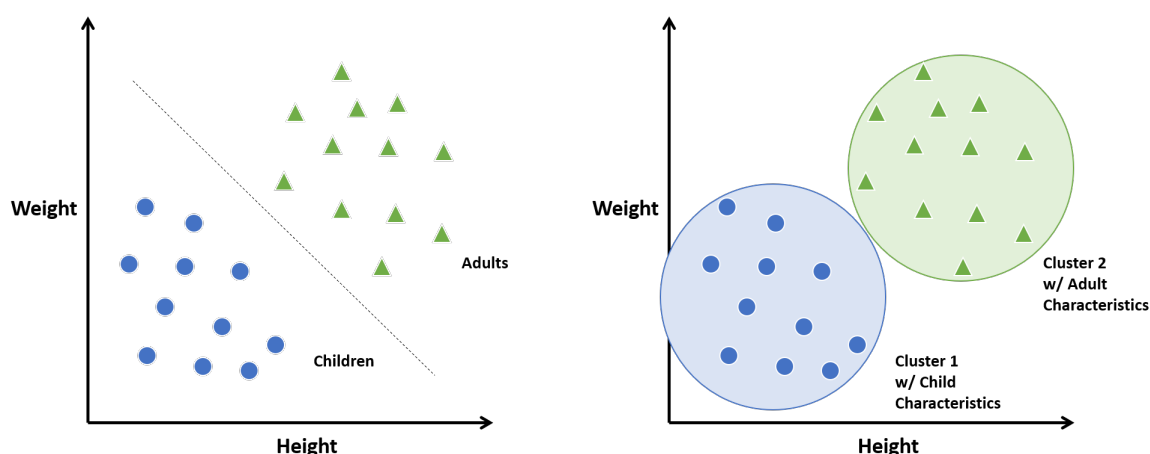
Essas relações hierárquicas voltadas à informação visam a facilitação da navegação através da organização, a recuperação das informações e o treinamento dos modelos de aprendizado em inteligência artificial.

6.1 Tipos de organização necessários no processo de clusterização

Clustering, ou clusterização, em seu sentido mais geral, refere-se ao processo de organizar um grupo de objetos com base em suas características compartilhadas (Nazari *et al.*, 2015) De acordo com os autores, é considerada uma tarefa de aprendizado indutivo, que envolve a extração de padrões ou estruturas dos dados sem nenhum conhecimento prévio do resultado desejado.

Ao contrário da classificação, a clusterização não visa prever um valor alvo predeterminado. Em vez disso, concentra-se na descoberta autônoma de classes ou *clusters* com base em padrões de similaridade encontrados nos dados. O principal objetivo da clusterização é organizar os dados que tenha alta similaridade em um mesmo cluster e de baixa similaridade entre diferentes *clusters*. O conceito de similaridade é frequentemente medido usando funções de distância, que de acordo com Bugatti (2008), consiste na medição da distância entre dois pontos em um espaço matemático, por exemplo, o espaço bidimensional cartesiano. Ela é empregada com o propósito de quantificar a similaridade entre conjuntos de dados.

Gráfico 2 - Diferença entre classificação e clusterização



Fonte: Lee (2020).

Classificação e agrupamento são dois métodos de identificação de padrões usados em aprendizado de máquina. A principal diferença entre eles está no fato de que a classificação utiliza classes predefinidas para as quais os objetos são atribuídos, enquanto o agrupamento identifica semelhanças entre objetos e os agrupa de acordo

com características comuns que os diferenciam de outros grupos de objetos (GORINI, c2024). Um exemplo para ilustrar essa diferença é em uma situação envolvendo espécies de animais. A classificação seria responsável por dizer que um grupo é de crianças e outro é de adultos. Já o agrupamento os organizaria por semelhança; por exemplo, ele agruparia por geração, por faixa etária e entre outras características pré-definidas pelo desenvolvedor do algoritmo.

Há vários requisitos que um bom algoritmo deve atender, como a escalabilidade, pois garante que o algoritmo possa lidar com grandes conjuntos de dados com eficiência. A escalabilidade é um conceito dentro da área de tecnologia da informação que busca prever o crescimento do micro e macro por meio de uma proporção, tem origem do termo em inglês, “*scalability*”, e é amplamente utilizada no contexto empresarial para permitir que a infraestrutura cresça em sintonia com o negócio. Almeida (2016) diz que a representação em escala é um instrumento para a compreensão micro e macro, e a sua usabilidade cresceu exponencialmente em tempos de Big Data.

Um dos desafios em relação à escalabilidade é a multiplicidade linguística. Este fator cria uma camada de complexidade a mais no processo de organização informacional devido às "armadilhas" associadas à linguagem humana. Algumas dessas armadilhas são explanadas por Alexopoulos (2020), em seu livro *Semantic Modeling for Data*, tais como: a ambiguidade, sinônimos e imprecisões. Ao implementar um sistema de governança eficaz, torna-se possível mitigar esses problemas por meio da estrutura taxonômica e, assim, assegurar que o conjunto de informações permaneça consistente, duradouro e relevante.

Os algoritmos de clusterização também devem ser capazes de sustentar o sistema apesar do ruído, pois, havendo erros devido a esse ruído, podem afetar os resultados do agrupamento. Além disso, a capacidade de aceitação dos dados dinâmicos é essencial ao lidar com *streaming* ou conjuntos de dados em evolução que mudam com o tempo.

Um aspecto adicional relevante é que a ordem de entrada dos dados não deve impactar o funcionamento do algoritmo. Isso significa que, independentemente da ordem em que os dados são inseridos, a saída do algoritmo deve ser sempre a mesma. Isso é crucial, especialmente quando lidamos com uma grande quantidade de dados de natureza diversa.

A eficiência de um algoritmo de clusterização é geralmente medida por sua capacidade de descobrir padrões ocultos nos dados. Um bom algoritmo deve ser capaz de identificar *clusters* independentemente de suas formas, além de detectar clusters que não sejam necessariamente esféricos ou bem separados.

No geral, o algoritmo de clusterização serve como uma ferramenta valiosa na análise de dados, permitindo a exploração e compreensão de conjuntos de dados complexos. Sendo assim, o processo de clusterização requer vários tipos de organização informacional para agrupar efetivamente os pontos de dados em agrupamentos significativos.

Os tipos mais comuns de organização informacional usados em sistemas de agrupamento são: agrupamento hierárquico, agrupamento *k-means* e agrupamento baseado em densidade.

Como foi citado anteriormente, a organização hierárquica é um tipo de ferramenta que usa uma estrutura semelhante a uma árvore para organizar os pontos de dados em *clusters*.

Para Nazari *et al.* (2015), os métodos hierárquicos são amplamente empregados em problemas de Mineração de Dados para fins de agrupamento.

Esses algoritmos fornecem uma abordagem sistemática para organizar os dados, permitindo uma melhor compreensão dos relacionamentos e padrões dentro deles. Entre os vários algoritmos hierárquicos, as abordagens de *bottom-up* são muitas vezes consideradas mais precisas em termos de agrupamento, mas apresentam uma compensação no custo.

Apesar da maior complexidade computacional, é importante observar que esse aumento de custo não implica necessariamente em aumento da complexidade conceitual ou algorítmica.

De fato, para os autores, o processo de formação da hierarquia do *cluster* pode ser dividido em uma série de operações básicas de fusão ou particionamento que podem ser executadas com eficiência.

Esse processo permite uma progressão clara e organizada na formação da hierarquia do *cluster*. Ao subdividir o método hierárquico, torna-se mais fácil compreender e implementar as diferentes etapas envolvidas, resultando em um processo de agrupamento mais gerenciável e eficiente.

Clusterização hierárquica aglomerativa: é uma abordagem de baixo para cima para clusterização, que visa agrupar objetos semelhantes. Nessa abordagem,

cada objeto inicialmente representa um *cluster* próprio. Em seguida, *clusters* semelhantes são mesclados interativamente até que a estrutura de *cluster* desejada seja obtida.

O algoritmo começa com N *clusters* - onde N é o número de amostras e cada *cluster* contém uma única amostra. O processo começa identificando os dois *clusters* com a semelhança mais próxima, com base em uma métrica de distância especificada. A métrica de distância pode ser definida usando vários critérios, como distância mínima, distância máxima, distância média ou distância central.

A escolha da métrica de distância depende dos requisitos específicos da tarefa de agrupamento. Depois que os dois *clusters* mais próximos são identificados, eles são mesclados em um único. Esse processo é repetido até que o número de *clusters* torne um ou o número especificado seja alcançado conforme determinado no projeto.

A clusterização hierárquica aglomerativa é versátil e amplamente utilizada, e pode ser aplicada a vários domínios e conjuntos de dados. Sua abordagem ascendente induz uma exploração flexível da estrutura de dados, permitindo a identificação de *clusters* significativos com base em sua similaridade.

Clusterização hierárquica divisiva: é uma abordagem de *top-down*, cujo objeto inicialmente pertence a um único *cluster* raiz e, em seguida, particiona os *clusters* existentes em *sub-clusters*. É o inverso do método aglomerativo.

Algoritmos de clusterização hierárquica ganharam interesse significativo em vários domínios de aplicativos devido à sua capacidade de produzir agrupamentos de alta qualidade.

Uma das vantagens é a flexibilidade na escolha de nível de granularidade. Esses algoritmos podem lidar com diferentes formas de similaridade ou medida de distância, tornando-os versáteis em cenários diversos. Além disso, podem ser aplicados a qualquer tipo de atributo, expandindo ainda mais sua função.

No entanto, há algumas desvantagens. A primeira delas é que a maioria dos algoritmos hierárquicos não revisita os *clusters* construídos depois de formados - isso significa que uma vez que um *cluster* é criado, ele não é refinado ou aprimorado.

A segunda desvantagem é que pode haver imprecisão nos critérios de término dos algoritmos de agrupamento hierárquico, dificultando a determinação de quando o processo de agrupamento deve parar.

Embora a maioria dos algoritmos de clusterização hierárquica envolva unir dois *sub-clusters* ou dividir um *cluster* em dois *sub-clusters* em uma única etapa, alguns algoritmos vão além disso e unem ou dividem mais de dois *sub-clusters*. No agrupamento hierárquico, o agrupamento de objetos pode ser representado como uma árvore chamada dendrograma.

No dendrograma elaborado por Nazari *et al.* (2015), os nós representam subconjuntos do conjunto de dados de entrada, com os nós-folha correspondendo aos elementos individuais do conjunto de dados e a raiz representando todo o conjunto de dados. Cada aresta no dendrograma representa uma relação de inclusão entre os clusters.

Assim, a clusterização hierárquica divisiva é uma abordagem de *top-down* que particiona os *clusters* existentes em *sub-clusters*. Ademais, oferece vantagens como flexibilidade na granularidade, facilidade de lidar com medidas de similaridade e distância e aplicabilidade a diferentes tipos de atributos.

No entanto, há limitações relacionadas a não revisitar *clusters* construídos para melhoria e imprecisão dos critérios de finalização. Na próxima figura, tem-se uma representação deste processo utilizando um dendrograma que fornece uma compreensão visual das relações de inclusão entre os agrupamentos.

Clusterização de particionamento: é a forma mais simples de clusterização, de acordo com os autores, em que um determinado conjunto de dados é dividido em "k-partes", com cada partição representando um *cluster*.

Ao contrário das abordagens hierárquicas, que dividem um *cluster* ou mesclam *sub-clusters*, os algoritmos de particionamento encontram todos os k-partes de uma só vez.

Esses algoritmos criam partições de um nível dos pontos de dados sem qualquer aninhamento. Para conseguir isso, a clusterização de particionamento utiliza esquemas heurísticos arrojados na forma de otimização iterativa. Esses esquemas envolvem diferentes estratégias de realocação, que reatribuem iterativamente pontos entre os k-partes, melhorando gradativamente o resultado do agrupamento.

Para uma clusterização obter sucesso usando esse método, cada *cluster* deve aderir a duas propriedades: conter pelo menos um objeto e cada objeto pertencer a exatamente um *cluster*.

Há vários métodos para clusterização de particionamento, incluindo *K-means*, *Bisecting k-means*, PAM (Partitioning Around Medoids) e o Clustering Probabilístico.

Dentre esses métodos, o *K-means* se destaca como um dos algoritmos de agrupamento mais populares e amplamente utilizados.

K-means é um algoritmo de *cluster* comumente usado em mineração de dados e reconhecimento de padrões. Tem como objetivo minimizar o índice de desempenho do *cluster*, usando erro quadrado e critério de erro.

O algoritmo começa selecionando os pontos focais iniciais do *cluster* e, em seguida, atribui os pontos de amostra restantes aos pontos focais mais próximos. Se a classificação não for razoável, é modificada, de modo a recalculer os pontos focais do *cluster*. Esse processo é repetido até que uma classificação aceitável seja obtida.

O algoritmo é conhecido por sua brevidade, eficiência e velocidade (Li; Wu, 2012). No entanto, depende muito dos pontos iniciais escolhidos, o que acarreta desfechos distintos. Além disso, o algoritmo pode ficar preso em pontos mínimos locais, caso o ponto focal inicial do cluster não seja apropriado.

Cada tipo de clusterização tem suas próprias vantagens e desvantagens. Selecionar o tipo apropriado de agrupamento para um determinado projeto é essencial para alcançar resultados precisos e significativos.

De modo a visar o aprimoramento dos sistemas de recomendação mais eficazes em plataformas de e-commerce, escolheu-se utilizar, neste projeto, a clusterização hierárquica, a óptica da taxonomia e o algoritmo de treinamento *k-means*.

Essas abordagens complementares fornecem uma forma sistemática de categorizar e organizar os dados, além de otimizar o processo de agrupamento. A taxonomia ajuda na criação de categorias estruturadas e significativas, enquanto o algoritmo *k-means* ajuda a encontrar os centróides de cluster ideais.

No entanto, observa-se que a escolha das técnicas de agrupamento deve ser baseada nos requisitos e características específicas do desenvolvimento da taxonomia. Crucial para o funcionamento, faz-se necessário métodos e abordagens específicas de construção, pois diferentes abordagens podem gerar resultados e desempenhos variados.

7 SISTEMAS DE RECOMENDAÇÃO

Um sistema de recomendação requer uma organização informacional que possa capturar de forma eficaz as preferências dos usuários e as características dos itens, conectar esses dados e recomendar um conteúdo/informação/produto baseado na similaridade entre as informações.

Como apontam Cazella, Nunes e Reategui (2010), esses sistemas operam reunindo recomendações de usuários e, em seguida, agregando-as e direcionando-as a indivíduos considerados potenciais destinatários de tais recomendações. Ao entender as relações entre usuários e itens, bem como os atributos desses itens, o sistema pode gerar recomendações personalizadas de acordo com o gosto de cada indivíduo.

Segundo a concepção de Ranganathan (2009) sobre "cada leitor tem seu próprio livro" gostaria de fazer uma analogia que pode ser adaptada ao contexto dos sistemas de recomendação. Neste cenário, cada recomendação desempenha um papel fundamental ao buscar atingir o seu leitor principal, visando a entrega de informações personalizadas para grupos específicos de usuários.

Por meio da análise de dados históricos, como compras anteriores, histórico de navegação e *feedback*, o sistema de recomendação pode identificar padrões e correlações que ajudem a determinar as preferências dos usuários. Esse sistema desempenha um papel crucial na facilitação de recomendações personalizadas.

Isso não apenas melhora a experiência do usuário, como aumenta as chances de os consumidores descobrirem novos itens que se alinhem a seus interesses. Ao alavancar técnicas avançadas de aprendizado de máquina e inteligência artificial, os sistemas de recomendação aprendem e refinam suas previsões, de modo a viabilizar uma experiência mais adequada ao contexto individual do usuário.

Um loop de feedback (também conhecido como *aprendizado de loop fechado*) descreve o processo de alavancar a saída de um sistema de IA e as ações correspondentes do usuário final para treinar novamente e melhorar os modelos ao longo do tempo. A saída gerada pela IA (previsões ou recomendações) é comparada com a decisão final (por exemplo, executar ou não o trabalho) e fornecer feedback ao modelo, permitindo que ele aprenda com seus erros (C3.AI, 2023, tradução nossa)

Além de aumentar a satisfação e o envolvimento do usuário em várias plataformas online, a recomendação assertiva de conteúdos como filmes, músicas,

livros ou produtos são fatores que desempenham um papel importante para a melhoria contínua durante o tempo de uso através do fornecimento dos *feedbacks*.

Conforme Schafer, Konstan e Riedl (1999), ao empregar os métodos de *feedbacks* nos sistemas de recomendação, ocorre uma contribuição para o aprimoramento da personalização, pois permite a adaptação de acordo com as necessidades e preferências exclusivas de cada cliente.

Além das interações usuário-item, a organização informacional também deve considerar o contexto em que as recomendações estão sendo feitas - fatores que podem ser exemplificados como a hora do dia, a localização do usuário e a atividade atual no navegador do usuário podem afetar significativamente a relevância e a utilidade das recomendações.

Portanto, é importante armazenar e processar dados relacionados aos contextos específicos. Ao incorporar a percepção de contexto no sistema de recomendação, é possível fornecer recomendações personalizadas que alinham com a situação específica do usuário.

7.1 Exemplos de uso

Compreendemos que o objetivo do sistema de recomendação é resultar na personalização dos produtos. A razão disso é que estamos lidando com produtos cada vez mais customizados, focados em atender as necessidades específicas de cada usuário e oferecer um serviço que possibilite uma melhor experiência.

Os sites de comércio eletrônico tornaram-se o foco principal da utilização de sistemas de recomendação, empregando várias técnicas para encontrar os produtos mais adequados para seus clientes e, finalmente, aumentar a lucratividade.

Esses sistemas utilizam várias técnicas para recomendar produtos, como a análise extensiva de dados, considerando a demografia do cliente ou até mesmo examinando o comportamento de compra para prever as preferências futuras.

Um exemplo notável pode ser observado em plataformas de streaming - como Spotify e Youtube. Essas plataformas revolucionaram a forma como consumimos entretenimento ao alavancar algoritmos sofisticados para analisar o comportamento e as preferências do usuário.

Ao observar o que os usuários assistem ou ouvem, esses sistemas podem fornecer recomendações precisas e relevantes, apresentando aos usuários novos

conteúdos que provavelmente irão gostar. Essa personalização transformou a maneira como descobrimos filmes, programas de TV e música, tornando a experiência mais agradável e adaptada aos nossos gostos individuais.

Além das plataformas de entretenimento, as recomendações personalizadas também são cruciais em outros domínios, como o e-commerce. Varejistas online, como a Amazon e Magazine Luiza, utilizam sistemas de recomendação para sugerir produtos com base no histórico de compras e navegação dos usuários.

Em vários domínios, como a esfera do entretenimento e comércio eletrônico, a capacidade de fornecer aos usuários recomendações pertinentes e personalizadas aumenta sua experiência geral, bem como desempenha um papel fundamental no reforço da trajetória triunfante e da aclamação generalizada das plataformas que empregam tais mecanismos.

Ao analisar detalhadamente as influências das sugestões nas empresas, podemos entender melhor os impactos profundos que elas têm, como mostram os exemplos a seguir.

Yahoo: um dos exemplos pioneiros desses sistemas é a empresa Yahoo. Cazella, Nunes e Reategui (2010), quando o My Yahoo introduziu o conceito de recomendação em larga escala, utilizando uma estratégia de personalização. Desde então, um número significativo de sites adotou esses sistemas de recomendação para fornecer aos usuários uma ampla gama de sugestões.

Youtube: o YouTube visa fornecer conteúdo relevante e envolvente para seus usuários, aumentando assim o tempo gasto na plataforma e a satisfação geral. Utiliza uma combinação de filtragem colaborativa e técnicas baseadas em conteúdo para sugerir vídeos personalizados aos usuários (Davidson *et al.*, 2010).

A filtragem colaborativa envolve a análise do comportamento do usuário, como histórico de visualizações e *likes/don't like*, para identificar padrões e realizar recomendações com base em preferências semelhantes do usuário. As abordagens baseadas em conteúdo, por outro lado, concentram-se na análise das características dos próprios vídeos, como *tags*, descrições e metadados para sugerir vídeos com conteúdo semelhante.

É importante mencionar uma possível "limitação" que gera um desafio no sistema de recomendação, que consiste na formação das bolhas de filtro, sendo este um dos desafios que o YouTube enfrenta - este tópico que será melhor abordado mais adiante no decorrer da pesquisa.

Estas bolhas surgem quando os usuários são expostos apenas ao conteúdo que se alinha somente com suas crenças e interesses existentes, levando a uma falta de diversificação na oferta e consumo de conteúdo (Davidson *et al.*, 2010). Essa bolha não se limita apenas a esta plataforma, mas também a outros serviços.

Apesar dos desafios, o YouTube continua a refinar seu sistema de recomendação para encontrar um equilíbrio entre personalização e diversidade, com o objetivo de fornecer aos usuários uma ampla variedade de conteúdo, maximizando seu envolvimento e satisfação.

Spotify: o Spotify, em particular, emergiu como uma força dominante na indústria da música, oferecendo aos usuários uma vasta biblioteca. O objetivo é selecionar músicas que os usuários gostariam de ouvir, na hora certa e no contexto certo.

Para atingir esse objetivo, o autor Pal (2021) afirma que o Spotify se concentrou em otimizar a organização do próprio conteúdo. Essa abordagem visa estimular que os usuários cliquem e reproduzam as listas de reprodução, bem como ajudá-los a descobrir novas músicas.

O princípio subjacente por trás dessa estratégia é semelhante ao aprendizado por reforço em que o Spotify aproveita os dados agregados do usuário para melhorar seu sistema de recomendação.

Ao analisar fatores tais como a quantidade de tempo que os usuários mantêm-se ouvindo as *playlists*, o sistema do Spotify pode diferenciar os aspectos únicos de consumo entre cada usuário e tratar a geração das *playlists* de forma distinta.

Se um indivíduo ouve frequentemente *playlists* e mixagens de jazz, o Spotify reconhece a singularidade desse comportamento de consumo e procura atendê-lo. A mesma abordagem pode ser encontrada no Youtube Music, Apple Music, Deezer e outros serviços de *audio streaming*. A plataforma reconhece que podem existir variações significativas no consumo e visa capturar essa diversidade como um aspecto essencial para aprimorar a experiência geral do usuário.

Por meio dessa abordagem, o Spotify continua a refinar seu algoritmo de recomendação, garantindo que os usuários recebam músicas que ressoam com as preferências pessoais de cada indivíduo, além de incentivá-los a explorar novos gêneros e artistas.

No entanto, essa conveniência tem um custo. Como Lüders (2019) argumenta, as recomendações algorítmicas e as listas de reprodução personalizadas

do Spotify podem desencorajar os usuários de ter um processo ativo em procurar e arquivar música.

Em vez disso, a plataforma incentiva os usuários a consumirem música passivamente com base em suas preferências estabelecidas. Isso levanta questões sobre o impacto do impulso de exploração do Spotify na inclinação dos usuários para arquivar música.

Embora não haja como negar que as recomendações algorítmicas do Spotify tornaram a música mais acessível, Lüders (2019) pontua que o Spotify também foi criticado por criar uma "bolha de filtro" musical.

Conseqüentemente, a vontade contínua dos usuários em consumir somente os mesmos conteúdos oferecidos pelo algoritmo pode estar em desacordo com o objetivo principal do próprio Spotify em promover a exploração musical e proporcionar ao usuário um comportamento ativo na plataforma.

Amazon: segundo estudo realizado por Hardesty (2019), o algoritmo de recomendação da Amazon passou por diversas transformações ao longo dos anos. Inicialmente, o algoritmo era baseado na filtragem colaborativa, que se baseava na análise do comportamento e preferências do usuário.

Essa abordagem provou ser eficaz em melhorar a satisfação do cliente e aumentar as vendas. No entanto, à medida que o volume de produtos na plataforma crescia, o algoritmo precisava se adaptar para lidar com o aumento da complexidade.

Para resolver isso, a Amazon introduziu uma abordagem híbrida que combinava filtragem colaborativa com filtragem baseada em conteúdo. Semelhante aos métodos adotados pelo Youtube, esse novo algoritmo levou em consideração as preferências do usuário e os atributos do produto, resultando em recomendações mais precisas. Além disso, a Amazon incorporou técnicas de aprendizado de máquina, como aprendizado profundo ou *deep learning* e processamento de linguagem natural (PLN) em seu sistema de recomendação, de modo a aprimorar ainda mais seu desempenho.

Ao alavancar grandes quantidades de dados e algoritmos avançados, o sistema de recomendação da Amazon tornou-se um componente crucial de seu modelo de negócios, contribuindo significativamente para seu sucesso como um dos maiores varejistas online do mundo.

Magazine Luiza: o sistema de recomendação do Magazine Luiza iniciou sua trajetória de expansão gradativa buscando se aprofundar em diversos domínios e fomentar o desenvolvimento de produtos inovadores.

De acordo com Félix, Tavares e Cavalcante (2018), Magazine Luiza reconheceu a importância de coletar e armazenar grandes quantidades de dados de várias fontes - como interações com clientes, transações de vendas e mídias sociais - para obter informações valiosas sobre o comportamento e as preferências do cliente.

O objetivo principal era estar em conformidade com os processos existentes e impulsionar o desempenho e a eficiência aprimorados. Ao adotar essa abordagem não convencional, a empresa conseguiu propiciar novas oportunidades e aproveitar melhor o potencial do sistema de recomendação, levando a um crescimento e sucesso no mercado brasileiro.

O investimento da empresa em tecnologias avançadas e sistemas de gerenciamento de dados ajudou a garantir processamento e análise de dados eficientes. Por sua vez, facilitou a melhor tomada de decisão e melhorou a experiência do cliente.

Em um mundo onde as recomendações personalizadas se tornaram a norma, é interessante visualizar exemplos da vida real em que as empresas aproveitaram o poder dos sistemas de recomendação.

De gigantes do comércio eletrônico, como a Amazon, a plataformas de *streaming*, como o Youtube, essas empresas revolucionaram a maneira como descobrimos produtos e conteúdo.

No entanto, à medida que nos aprofundamos no mundo dos sistemas de recomendação, é crucial esclarecer uma outra perspectiva já citada anteriormente: as possíveis desvantagens escondidas sob a superfície.

7.2 Prós e contras dos sistemas de recomendação

Não obstante, há preocupações quanto às desvantagens dos sistemas de recomendação, principalmente relacionadas a bolhas de filtro e câmaras de eco. De acordo com Davidson *et al.* (2010), as bolhas de filtro referem-se ao fenômeno em que os usuários são expostos apenas ao conteúdo que se alinha com suas crenças e preferências existentes. Isso pode potencialmente limitar sua exposição a diversas perspectivas e sufocar a troca de ideias diferentes.

Por outro lado, as câmaras de eco ocorrem quando crenças e opiniões existentes são reforçadas dentro de um ambiente online fechado. Tanto as bolhas de filtro quanto as câmaras de eco podem contribuir para a polarização dos pontos de vista, bem como dificultar o desenvolvimento de perspectivas completas.

Câmaras de eco ocorrem quando pessoas com os mesmos interesses ou pontos de vista interagem principalmente dentro de seu grupo. Eles buscam e compartilham informações que estão em conformidade com as normas de seu grupo e tendem a reforçar as crenças existentes (Jamieson & Cappella, 2008; Sunstein, 2009). O fato de a psicologia social ter mostrado há muito tempo essa tendência de associar-se com outros que pensam da mesma forma é comum em várias culturas. No entanto, há um novo medo de que o atual sistema de mídia esteja ajudando as pessoas a entrar nas câmaras de eco com mais facilidade do que nunca (Dubois; Blank, 2018, p. 731, tradução nossa).

Em ambas as situações, os usuários são expostos a uma variedade restrita de artistas, gêneros e conteúdo em geral, consumindo apenas um tipo de informação e até mesmo proveniente da mesma fonte, algo que ocorre frequentemente durante períodos eleitorais. Esse fenômeno pode ser observado em plataformas de *streaming* de áudio, *streaming* de vídeo e até nos *feeds* das redes sociais, abrangendo diferentes tipos de plataformas.

Apesar dessas desvantagens, ter um produto que possa entender as necessidades dos usuários e recomendar os produtos e informações mais adequados a eles pode melhorar significativamente a experiência do usuário.

Esse nível de personalização aumenta a conveniência do usuário e possibilita que sejam expostos a uma gama diversificada de opções que, de outra forma, poderiam não ter acesso.

Quando os usuários sentem que a plataforma entende suas preferências e entrega consistentemente conteúdo relevante, é mais provável que continuem usando o serviço e o recomendem a outras pessoas devido a boa experiência.

Dessa forma, pode-se levar ao aumento da satisfação e lealdade do usuário, tornando provável que os usuários se convertam a *heavy users* e até mesmo promotores dos serviços oferecidos por essas empresas.

Constando esses aspectos, é perceptível que um componente importante dos sistemas de recomendação é a organização da informação através do agrupamento de informações por semelhança, porque ajuda a identificar grupos de itens ou usuários similares.

Ao agrupar itens ou usuários, o sistema pode fazer recomendações mais precisas, concentrando-se em itens ou usuários semelhantes àqueles com os quais o usuário já interagiu.

Além disso, o agrupamento pode ajudar a identificar tendências no comportamento do usuário, permitindo que o sistema faça previsões mais precisas sobre quais itens esse usuário pode estar interessado.

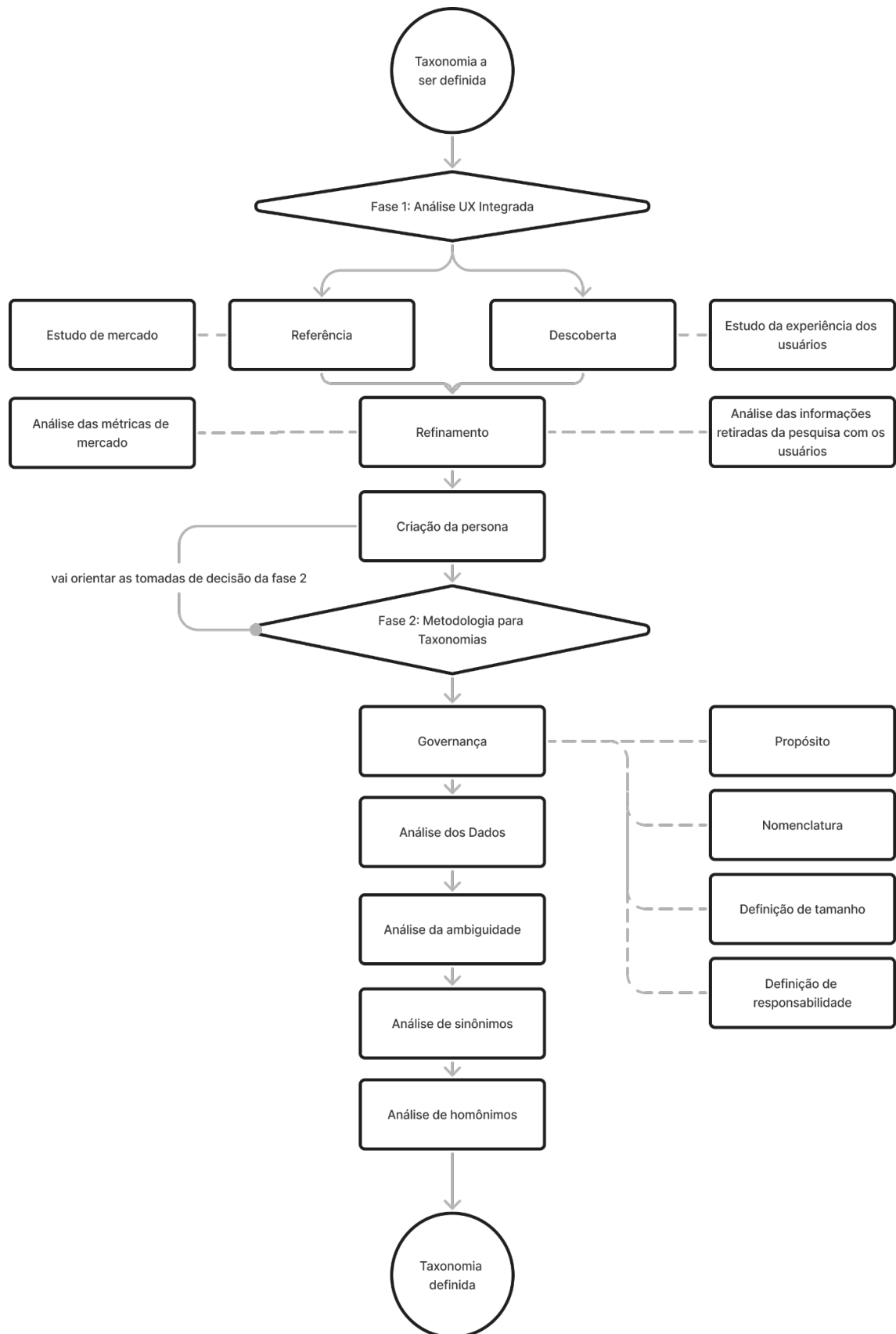
8 DESENVOLVIMENTO DA METODOLOGIA DE CONSTRUÇÃO DE TAXONOMIA

O desenvolvimento deste estudo começou com a definição clara do objetivo de criar uma proposta de uma metodologia de construção de taxonomia. A taxonomia é uma ferramenta essencial para a classificação e organização de informações, criando hierarquias que facilitam o entendimento e a utilização dos dados. Esta ferramenta de conhecimento técnico pode assumir diversos formatos, sendo crucial alinhar pontos principais para garantir o melhor desenvolvimento.

8.1 Uma metodologia de construção de taxonomia

A metodologia proposta inicia-se na Fase 1: Análise UX Integrada, que adota uma abordagem qualitativa e quantitativa. Nessa fase, é analisado o contexto em que o usuário está inserido, o mercado no qual ele consome e como o mercado se posiciona. O objetivo é obter insumos suficientes para embasar a construção da taxonomia e potencializar seu uso, ampliando a possibilidade de êxito da proposta. Em seguida, tem início a Fase 2: Metodologia para Taxonomias, que tem foco exclusivo na análise e criação dos elementos essenciais para a estrutura taxonômica cumprir seu objetivo.

Figura 1 - Fluxo das fases e etapas



Fonte: Elaborado pela autora (2024).

A figura acima ilustra o método proposto, facilitando a compreensão geral. Ela apresenta um fluxo de cada etapa da metodologia para a construção de taxonomias, oferecendo um panorama geral do processo. No entanto, antes de detalhar cada etapa do método, é necessário discutir como ele se diferencia dos métodos atuais existentes, bem como identificar e relacionar quais são esses métodos.

Ao buscar a melhor forma de construir um método para desenvolver taxonomias, foram examinados trabalhos anteriores e, durante a revisão de literatura, a diversidade de cenários em que a taxonomia era utilizada como ferramenta foram considerados, além dos diferentes métodos de criação de taxonomia para contextos específicos. Notou-se uma variedade de abordagens, desde métodos voltados ao planejamento ambiental, como o de Barbosa *et al.* (2022), até a construção de aparelhos elétricos, como descrito por Lam, Fung e Lee (2007).

Esses estudos exploraram diversas abordagens e metodologias, analisando como cada uma poderia aplicar a taxonomia de forma eficaz em seus respectivos contextos. No entanto, considerando-se a área de aplicação e estudo da presente pesquisa, foram considerados apenas aqueles que eram voltados ao ambiente informacional e sua aplicabilidade no universo da tecnologia. Entre os métodos destacados, estão o método Delphi (Strasser, 2017), o método de Bayona-Oré *et al.* (Usman *et al.*, 2017), o método de Construção Automática de Taxonomia de Domínio a Partir de Texto - ADTCT (Knijff; Frasinca; Hogenboom, 2013) e o método para desenvolvimento de taxonomia e sua aplicação em sistemas de informação (Nickerson; Varshney; Muntermann, 2013).

O método Delphi (Strasser, 2017) é utilizado para examinar problemas complexos com a ajuda de um grupo de especialistas selecionados devido ao seu conhecimento especializado. Ele se caracteriza pelo feedback controlado, processo iterativo, agregação estatística e anonimato dos participantes.

O processo de uma taxonomia baseada no método Delphi envolve as seguintes etapas:

- Escolha da meta-característica do objeto de interesse.
- Especificação das dimensões.
- Definição das condições necessárias da taxonomia.
- Conceitualização das características da taxonomia.

- Através dessas etapas, realiza-se uma análise cruzada entre as dimensões e características da taxonomia, proporcionando uma perspectiva ampla e especializada com foco e objetivo bem definidos, tudo isso mantendo o anonimato dos participantes.

O método Delphi fornece uma visão clara dos cenários e etapas da taxonomia, o que gera uma gama de alternativas e decisões a serem tomadas. No entanto, ele não especifica como abordar cada etapa, deixando isso em aberto e subjetivo conforme a necessidade de cada situação. Embora isso seja válido para uma taxonomia, busquei em minha proposta reduzir essa subjetividade e trazer uma generalidade que permita a utilização do método em diferentes cenários.

O método de Bayona-Oré *et al.* (Usman *et al.*, 2017) é uma proposta para o desenvolvimento de taxonomias especificamente utilizada para criar uma taxonomia de fatores críticos de sucesso na implantação de processos de software. Esse método organiza atividades em cinco grupos, considerados as fases do método proposto:

- Planejamento: estabelece planos para definir as atividades para o desenho e implementação da taxonomia.
- Identificação e Extração: alinha o plano de trabalho com as informações exigidas pela organização.
- Projeto e Construção: desenha e constrói a taxonomia utilizando o inventário de termos para determinar sua estrutura final.
- Teste e Validação: garante que a taxonomia desenhada será útil para os usuários atingirem seus objetivos.
- Implantação: implementa a taxonomia em toda a organização.

Esse método é particularmente interessante para criar uma taxonomia a partir de um inventário de termos, o que é inovador, especialmente nas fases iniciais, quando não se tem conhecimento dos termos presentes na base de dados. A taxonomia pode servir como uma ferramenta para identificar e extrair termos, ajudando a construir uma estrutura taxonômica básica para um funcionamento imediato e oferecendo soluções práticas.

A fase de Projeto e Construção, que utiliza o inventário de termos para determinar a estrutura final da taxonomia, é especialmente inovadora. Essa

abordagem pode ser uma alavanca para sistemas em geral, facilitando a identificação e extração de termos e ajudando a construir uma taxonomia base para funcionamento imediato e soluções práticas.

O método ADTCT (Knijff; Frasinca; Hogenboom, 2013) utiliza um analisador sintático (part-of-speech parser) para extrair termos de um corpus de texto. Os substantivos extraídos são filtrados com base em critérios, como: pertinência ao domínio, consenso de domínio, coesão lexical e relevância estrutural - deixando apenas os conceitos de domínio.

Para organizar hierarquicamente esses conceitos em classes de taxonomia, são utilizados dois métodos:

- Método de Subsunção: baseia-se na coocorrência de documentos. Um conceito subsumi outro se a ocorrência do segundo conceito está sempre associada à ocorrência do primeiro conceito, mas não o contrário. Esse método rotula e organiza os conceitos em uma hierarquia, resultando em uma taxonomia que, segundo os autores, é continuamente rasa. Uma taxonomia mais profunda teria mais informações sobre as relações taxonômicas, mas essas relações seriam menos confiáveis.
- Algoritmo de Agrupamento Hierárquico: utiliza a similaridade de coocorrência como medida de distância. Isso significa que a coocorrência de conceitos em diferentes documentos é analisada. Para cada conceito, os possíveis "pai-conceitos" e seus subconceitos são determinados por sua relação hierárquica.

Esse método é muito interessante, especialmente porque permite a análise de textos e conceitos de forma ampla, possibilitando a escalabilidade. No entanto, como apontado pelos próprios autores, os resultados tendem a ser superficiais, exigindo continuidade de treinamento, experimentação e reforço por parte dos integrantes para alcançar um resultado satisfatório. Muitas vezes, é necessária a intervenção de especialistas e a utilização de uma taxonomia como base para comparação e aprendizado em relação à sua estrutura. Pode ser um método valioso para estudar a similaridade entre categorias da estrutura taxonômica e analisar aspectos como governança.

Nickerson, Varshney e Muntermann (2013) propõem um método abrangente para o desenvolvimento de taxonomias, detalhando o processo completo de construção e avaliação dentro do ciclo da ciência do design. O método visa garantir a utilidade e aplicabilidade das taxonomias desenvolvidas, submetendo-as a um conjunto de condições necessárias.

Os autores destacam que a maior contribuição desse método é a metacaracterística, que se baseia no propósito da taxonomia e no uso esperado pelos usuários. Esse ponto converge com a minha proposta, onde também identifico o propósito como o primeiro item a ser analisado, referindo-me a ele como governança, enquanto os autores o denominam metacaracterística.

As etapas do método são:

- Determinação da metacaracterística: definir a característica principal da taxonomia com base no propósito e nos usuários.
- Determinação das condições finais: estabelecer as condições que a taxonomia deve cumprir.

Empírico para conceitual:

- Identificar (novo) subconjunto de objetos: selecionar objetos com os quais o pesquisador está mais familiarizado ou que são mais facilmente acessíveis, possivelmente através de uma revisão da literatura.
- Identificar características comuns e agrupar objetos: as características devem ser consequências lógicas da metacaracterística.
- Agrupar características em dimensões para criar (ou revisar) a taxonomia: agrupar formalmente usando técnicas estatísticas ou informalmente usando processos manuais ou gráficos.

Conceitual para empírico:

- Conceitualizar (novas) características e dimensões dos objetos: dedutivamente, com base nas noções do pesquisador sobre como os objetos são semelhantes e diferentes.
- Examinar os objetos quanto às características e dimensões: verificar se as dimensões contêm características que são mutuamente exclusivas e coletivamente exaustivas.

- Criar (ou revisar) a taxonomia: desenvolver ou ajustar a estrutura taxonômica com base nas análises anteriores.

De todas as abordagens, essa é a que mais se alinha com o tópico de criação de taxonomia e governança proposto neste trabalho, já que ambas buscam definir o propósito da taxonomia. Além disso, a etapa relacionada à identificação de características e objetos é semelhante à proposta desta pesquisa, no que concerne a análise de dados. No entanto, os autores não abordam a análise de dados em profundidade, nem consideram questões como ambiguidade, sinônimos e homônimos - ao contrário da proposta desta pesquisa. Não obstante, vale ressaltar que os autores trouxeram etapas de design para dentro do processo de taxonomia, enquanto na proposta desta pesquisa a abordagem analisa esse tópico anterior de forma complementar.

Em geral, o método de Nickerson, Varshney e Muntermann (2013) é extremamente interessante, especialmente para estabelecer uma boa governança de taxonomia, um processo robusto de desenvolvimento e um fluxo de trabalho com pouca interferência externa.

Após analisar os métodos anteriores, percebe-se que não há uma única forma correta de construir uma taxonomia. A diversidade de cenários e aplicabilidades torna difícil chegar a um consenso sobre as melhores práticas para a criação de uma taxonomia. O que funciona em um contexto pode não ser eficaz em outro.

Embora haja convergências entre os métodos dos autores analisados, a proposta da presente pesquisa visa complementar esses métodos com uma abordagem adicional. Essa abordagem é projetada para ser aplicada em diferentes cenários, mas com um enfoque maior na potencialização dos sistemas de recomendação.

Os diferenciais da presente proposta incluem:

- **Design de experiência como fase anterior à construção da taxonomia:** o design de experiência é considerado uma etapa preliminar essencial, capaz de fornecer insumos valiosos para a criação da taxonomia.
- **Persona:** resultado de todo o processo de design, a persona é um elemento diferencial para guiar todas as fases de criação das taxonomias, servindo como um direcionador fundamental para todo o processo.

- **Governança, tamanho, ambiguidade, sinônimos e homônimos:** a definição e análise desses elementos são incorporadas como parte integral do processo de construção da taxonomia.

Essa abordagem busca oferecer um método mais abrangente e flexível, capaz de ser adaptado a diversos contextos, mas com enfoque na eficácia dos sistemas de recomendação. Para uma melhor compreensão, ambas as fases são explicadas com detalhe nas próximas seções.

8.1.1 Fase 1: Análise UX Integrada

O método utilizado na construção de taxonomias, como se vê na Figura 1, está dividido da seguinte maneira na Fase 1:

- **Descoberta:** a descoberta permite tomar decisões cruciais que realmente atendem às necessidades do usuário, em vez de se basear apenas em suposições. Compreender questões específicas é fundamental para estabelecer uma base sólida para o desenvolvimento do projeto, bem como para entender as necessidades, frustrações e comportamentos dos usuários.
 - a. **Estudo da experiência dos usuários:** o foco aqui será estudar a experiência do usuário. Foi analisado os pontos de interação que os usuários têm com o sistema de recomendação, quais dispositivos utilizam, a forma como as informações estão organizadas e todos os outros elementos que compõem a interação entre o sistema e o usuário. Além disso, será avaliada a experiência completa do usuário ao utilizar esses sistemas e a sua perspectiva em relação aos resultados proporcionados por essa experiência.
- **Referência:** é um processo estratégico que permite medir ou avaliar seu produto ou estudo em relação a outras iniciativas do mercado. Por meio desse processo, é possível identificar oportunidades e entender a concorrência, visando a melhoria contínua e garantindo que sua oferta esteja alinhada com as tendências do mercado, evitando a produção de algo que não seja comercializável.
 - a. **Estudo de mercado:** o foco aqui foi entender o mercado, identificando as oportunidades que ele apresenta em termos de setores, faixa etária, tecnologia e informações demográficas sobre o consumidor brasileiro.

Esse entendimento proporciona insumos suficientes para, na fase de refinamento, comparar os dados obtidos na pesquisa do usuário. Dessa forma, é possível verificar quais informações estão em sintonia, quais diferem e quais lacunas de mercado podem ser preenchidas com dados fornecidos pelos usuários, agregando valor ao projeto final.

- Refinamento: este é um momento crucial de análise para extrair ideias interessantes e cruzar informações das fases anteriores. É fundamental compreender as necessidades e frustrações dos usuários, assim como as inovações e métricas que o mercado oferece. Analisar e cruzar essas pesquisas proporciona insumos valiosos que orientam toda a estrutura organizacional da taxonomia, garantindo sua usabilidade e eficácia no produto final.
 - a. Análise das informações retiradas da pesquisa com o usuário: aqui, a análise é voltada para a jornada do usuário, incluindo sua primeira impressão, as dores e necessidades atendidas, bem como a interatividade com os produtos e sistemas.
 - b. Análise das métricas de mercado: análise de gráficos, tendências e comportamento do mercado.
- Criação da persona: ao mergulhar na experiência do usuário, informações valiosas podem ser obtidas em relação às necessidades, preferências e comportamentos específicos de indivíduos que se envolveriam diretamente com a taxonomia. Esta investigação empírica esclarecerá as características, expectativas e pontos problemáticos dos beneficiários pretendidos, permitindo uma abordagem mais personalizada e eficaz. Uma exploração do cenário do mercado fornecerá uma compreensão abrangente do contexto mais amplo, no qual a taxonomia irá operar. Esse exame contemplará fatores como tendências do setor, dinâmica competitiva e demanda do mercado, todos os quais influenciam a receptividade e a viabilidade da taxonomia. A persona tem como propósito a generalização do perfil dos entrevistados representado por uma pessoa fictícia. É uma representação baseada nos dados retirados da pesquisa, complementada por algumas tendências de mercado.

8.1.2 Fase 2: Metodologia para Taxonomias

O método utilizado na construção de taxonomias, como se vê na Figura 1, está dividido da seguinte maneira na Fase 2:

- Governança: para garantir a pertinência e eficácia contínuas das taxonomias, Wells (2018) afirma a necessidade de manutenção e adaptação regulares. É essencial que as taxonomias evoluam para acomodar novos tipos de materiais e mudanças no cenário social ou organizacional. Para estabelecer a governança é preciso ter a persona, o propósito claro da taxonomia, definição de tamanho e de responsabilidade, veja a seguir.
 - a. Propósito: delimitar o objetivo da taxonomia e sua utilidade pretendida.
 - b. Nomenclatura: o nome dado a cada rotulagem/etiqueta, ou mais conhecido na Biblioteconomia, classe, dentro da taxonomia.
 - c. Definição de tamanho: ao examinar o nível de granularidade dentro de uma classe hierárquica, é crucial fazer uma análise abrangente. Isso implica avaliar o grau de granularidade presente em cada nível da estrutura hierárquica. Por exemplo, uma determinada taxonomia necessita de um mínimo de quatro níveis de granularidade para cumprir efetivamente seu propósito. Em contraste, outra taxonomia requer uma estrutura mais intrincada, exigindo, por exemplo, até seis níveis de granularidade. Ao investigar e comparar estes níveis de granularidade, torna-se possível obter informações valiosas sobre as complexidades e variações das classificações hierárquicas. Esta abordagem analítica permite uma compreensão mais robusta das estruturas hierárquicas e sua adequação para diferentes contextos e aplicações.

Em resumo: avaliação do nível de granularidade de cada nível hierárquico, com exemplos variando de quatro a seis níveis de granularidade, conforme a necessidade da taxonomia.
 - d. Definição de responsabilidade: o papel do taxonomista e a alocação desse profissional dentro da estrutura organizacional são de grande importância em relação à curadoria e manutenção da taxonomia. O taxonomista, como membro de uma equipe, ou das equipes, desempenha um papel fundamental no estabelecimento e organização da taxonomia.

- **Análise dos Dados:** torna-se imperativo fazer seleções meticulosas sobre os dados que são comumente usados e têm influência significativa na semântica da taxonomia, a fim de contornar possíveis problemáticas. Algumas dessas problemáticas são a ambiguidade, homônimo e sinônimo.
- **Ambiguidade:** refere-se à ausência de clareza, levando à presença de interpretações duais ou significados incertos. Por exemplo, a afirmação “não sei se gosto do frio ou do calor” exemplifica o dilema das preferências incertas entre temperaturas contrastantes. A ausência de clareza leva a múltiplas interpretações.
- **Sinônimos:** Bechara (2009) afirma que sinônimos são palavras com semelhante significação, podendo uma estar no lugar da outra em determinado contexto. Por exemplo, as palavras "belo" e "lindo", que compartilham atributos sinônimos, possuem grafias diferentes, mas com semelhança no significado. Sinônimos, portanto, servem como ferramentas linguísticas que facilitam a expressão de ideias com vocabulário variado, permitindo que os indivíduos acrescentem profundidade e nuances à sua comunicação. Ao empregar sinônimos, escritores e falantes podem efetivamente transmitir de diferentes formas o significado pretendido, habilidade que uma máquina desconhece.
- **Homônimos:** de acordo com Bechara (2009), homonímia é a característica de duas ou mais palavras, com significados ou funções diferentes, possuírem a mesma estrutura fonológica (mesmos sons e acentuação). Determinar a distinção entre homônimos é crucial para a compreensão linguística e comunicação precisa. Uma ilustração pertinente desse conceito é a palavra "grama", que pode denotar tanto um vegetal quanto uma unidade de medida. Assim, os homônimos servem como um fascinante objeto de investigação no estudo da linguagem, pois desafiam os usuários a navegar pelas nuances de múltiplos significados associados a itens lexicais aparentemente idênticos. Em resumo: palavras com pronúncia e ortografia idênticas, mas significados diferentes.

8.2 Experimentação e Validação

Será demonstrado a aplicação do método usando as etapas, entradas e saídas apontadas nas fases e etapas. A seguir, o quadro com etapas e definições.

Quadro 2 - Etapas e respectivas definições

Etapas	Definição
Descoberta	Estudo da experiência dos usuários: entender a perspectiva do usuário sobre a eficácia da experiência com os elementos da interação com o sistema de recomendação.
Referência	Estudo de mercado: entender o mercado identificando as oportunidades.
Refinamento	<p>Análise das informações retiradas da pesquisa com o usuário: a análise é voltada para a jornada do usuário.</p> <p>Análise das métricas de mercado: análise de gráficos, tendências e comportamento do mercado.</p>
Criação da persona	Generalização do perfil dos entrevistados representado por uma pessoa fictícia.
Governança	<p>Propósito: delimitar o objetivo da taxonomia e sua utilidade pretendida.</p> <p>Nomenclatura: o nome dado a cada rotulagem/etiqueta ou classe dentro da taxonomia.</p>

	<p>Definição de tamanho: o nível de granularidade dentro de uma classe hierárquica</p> <p>Definição de responsabilidade: o papel do taxonomista dentro de todo o processo.</p>
Análise dos Dados	Seleções meticulosas sobre os dados que são comumente usados e têm influência significativa na semântica da taxonomia
Ambiguidade	Análise de termos e classes de interpretações duais ou significados incertos
Sinônimos	Análise de termos e classes com significados comparáveis, embora diferentes na forma.
Homônimos	Análise de palavras com pronúncia e ortografia idênticas, mas significados diferentes

Fonte: Elaborado pela autora (2024).

Quadro 3 - Entradas e saídas de cada etapa

Etapas	Entradas	Saídas
Estudo da experiência dos usuários	Entrevista com o usuário.	Documento que explicita a entrevista na íntegra.

Estudo de mercado	Pesquisa de fontes confiáveis para obter informações e ideias sobre o mercado.	Gráficos, tabelas, métricas de tendências
Análise das informações retiradas da pesquisa com o usuário	Gráficos, tabelas, métricas de tendências obtidas da pesquisa com o usuário	Documento que explicita a perspectiva geral do usuário, suas dores e necessidades atendidas, seus desejos, a compreensão completa de sua jornada com o sistema e apresenta dados demográficos para análise da amostra.
Análise das métricas de mercado	Citações, gráficos e tabelas obtidas do estudo de mercado	Documento com números percentuais comparativos, geração de novas ideias, identificação de tendências e de lacunas que a pesquisa do usuário pode ter deixado em aberto
Criação da persona	Documentos das etapas: análise das informações retiradas da pesquisa com o usuário e análise das métricas de mercado	Imagem representativa da persona
Propósito	Identificação do objetivo utilizando a persona	Documentação oficial de taxonomia que contém o objetivo principal a ser atingido, persona e informações resumidas que serão atualizadas conforme o avanço das etapas do desenvolvimento da

		taxonomia até a definição de responsabilidades. Pode também sair um fluxograma.
Nomenclatura	Imagem representativa da persona	Documentação sobre a linguagem ideal a ser utilizada, baseada nas características e preferências da persona.
Definição de tamanho	Definição do nível de granularidade dentro de uma classe hierárquica	Documentação de quantas classes hierárquicas existirão na taxonomia.
Definição de responsabilidade	Definição das responsabilidades do taxonomista e outros colaboradores do projeto	Documentação da definição dos papéis e suas respectivas funções em cada etapa do processo.
Análise dos Dados	Tabelas a serem consultadas em sistemas e arquivos em Excel.	Documentação ou Excel com a lista de termos para análise de sinônimos, homônimos e ambiguidade.
Ambiguidade	Documentação ou Excel com a lista de termos para análise	Documentação ou planilha Excel contendo a lista de termos ambíguos cuidadosamente evitados.

Sinônimos	Documentação ou Excel com a lista de termos para análise	Documentação ou planilha Excel contendo a lista de sinônimos cuidadosamente selecionados.
Homônimos	Documentação ou Excel com a lista de termos para análise	Documentação ou planilha Excel contendo a lista de termos homônimos cuidadosamente evitados.

Fonte: Elaborado pela autora (2024).

Os dois quadros apresentados servem como suporte visual. O Quadro 2 facilita a visualização das etapas e suas respectivas definições, estabelecendo um conceito claro e resumido de cada fase do processo, evitando confusões de conceito durante a aplicação da metodologia.

O Quadro 3 apresenta as entradas e saídas de cada etapa. As entradas são os elementos utilizados para iniciar cada fase, enquanto as saídas indicam o resultado ou documento final de cada uma delas. Dessa forma, ficam claras as expectativas de resultados e os documentos gerados em cada etapa, o que é fundamental para manter o processo organizado e garantir uma expectativa clara de resultados.

8.2.1 Fase 1: Análise UX Integrada

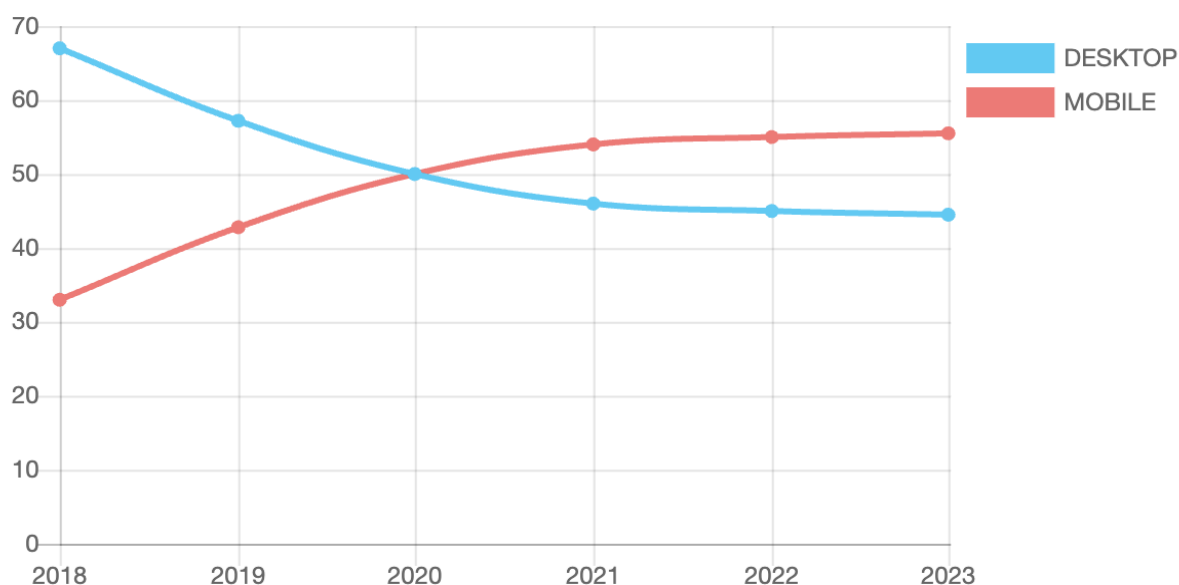
Concentrei-me no mercado de e-commerce devido a análise da Neiva (2024), que utilizou dados da Associação Brasileira de Comércio Eletrônico (Abcomm). A previsão é que o faturamento do e-commerce no Brasil continue crescendo, podendo ultrapassar os R\$ 250 bilhões até 2027, o que motiva a criação constante de lojas virtuais. Nesse contexto, entender o consumidor é essencial para se destacar da concorrência. Em 2023, o faturamento do e-commerce no Brasil foi de R\$ 185,7 bilhões.

Neiva (2024) também revela que o consumidor brasileiro está cada vez mais digital. Dados da pesquisa E-commerce Trends, realizada em parceria entre Opinion

Box e Octadesk, com 2.088 consumidores brasileiros acima de 16 anos, em maio de 2023, mostram que 58% dos entrevistados compraram mais online do que fisicamente em 2023, enquanto 22% ainda preferem lojas físicas.

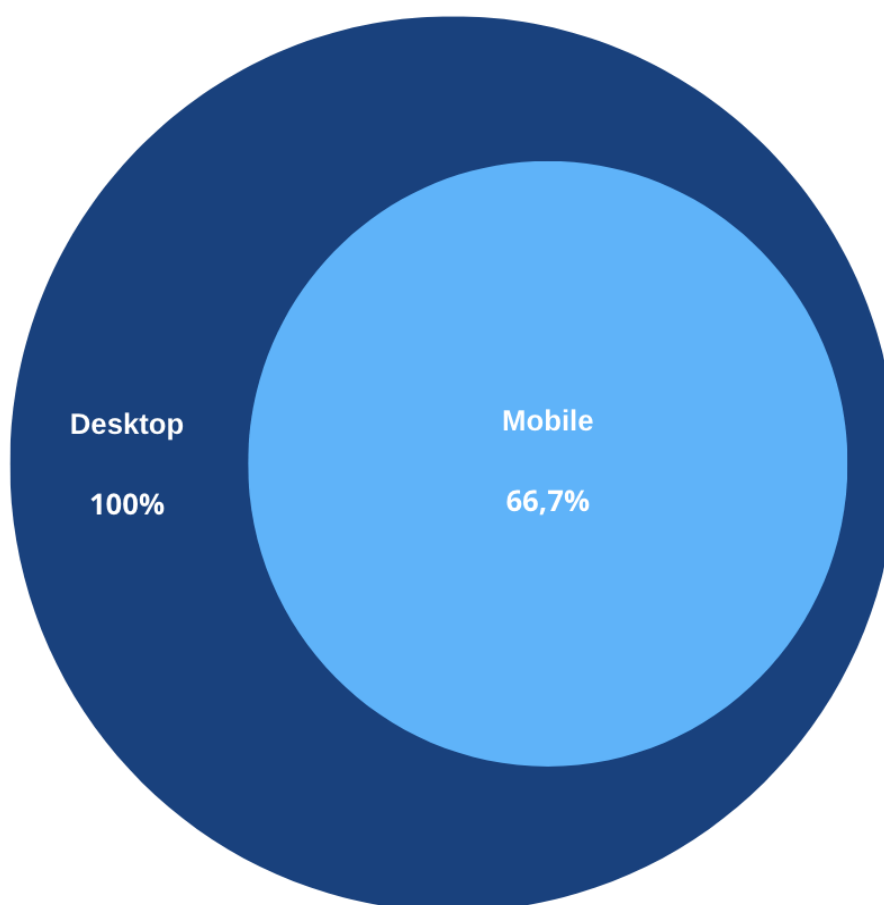
Adicionalmente, a Abcomm e Neiva (2024) indicam que, em 2024, as mulheres representam 59% dos clientes de e-commerce, e mais de 50% dos consumidores virtuais têm entre 25 e 44 anos, estando concentrados na região Sudeste do país. Fazendo um comparativo, podemos afirmar que a entrevista segue a linha de raciocínio analisada por Neiva (2024) e está dentro dos cenários apontados pela Abcomm. Através da pesquisa feita com os usuários, nove pessoas participaram, com idades entre 24 e 38 anos. A amostra inclui três pessoas do Nordeste, uma do Sul e cinco do Sudeste, sendo São Paulo o Estado com mais participantes. A divisão de gênero é de 44,4% masculino e 55,6% feminino.

Gráfico 3 - Dispositivos utilizado na compra dos brasileiros de acordo com Abcomm



Fonte: Abcomm (2024).

Gráfico 4 - Dispositivo de preferência para fazer compras online



Fonte: Elaborado pela autora (2024).

A análise da entrevista também levou em conta os dispositivos utilizados para compras, conforme os dados apontados pela Abcomm, que mostram que o consumo via mobile representa cerca de 55% de todas as vendas, comparado a 45% via desktop, e essa proporção tende a aumentar nos próximos anos.

Entre os nove entrevistados, 100% (nove pessoas) afirmaram utilizar o notebook para navegar nas ofertas e finalizar as compras, ao mesmo tempo que 66,6% (seis pessoas) relataram que também utilizam o celular paralelamente.

Podemos afirmar que o público-alvo da nossa pesquisa é predominantemente feminino, com foco na região Sudeste e com alta utilização de dispositivos desktop. É interessante, pois ao contrário do que foi observado pela Abcomm, o dispositivo mais utilizado para compras nos dados fornecido por eles é o mobile.

Trouxe alguns trechos fornecidos pelos usuários, aos quais chamaremos, respectivamente, de “Resposta 1, 2 e 3”, para justificar a preferência por um tipo de dispositivo ou mais, veja a seguir:

A facilidade de uso pelo celular é a maior vantagem. Para produtos como itens de casa, roupas ou produtos para pets, prefiro usar o computador para ter uma melhor visualização. Para consumo imediato, como comida, uso aplicativos no celular. A praticidade é essencial na minha rotina. (Resposta 1).

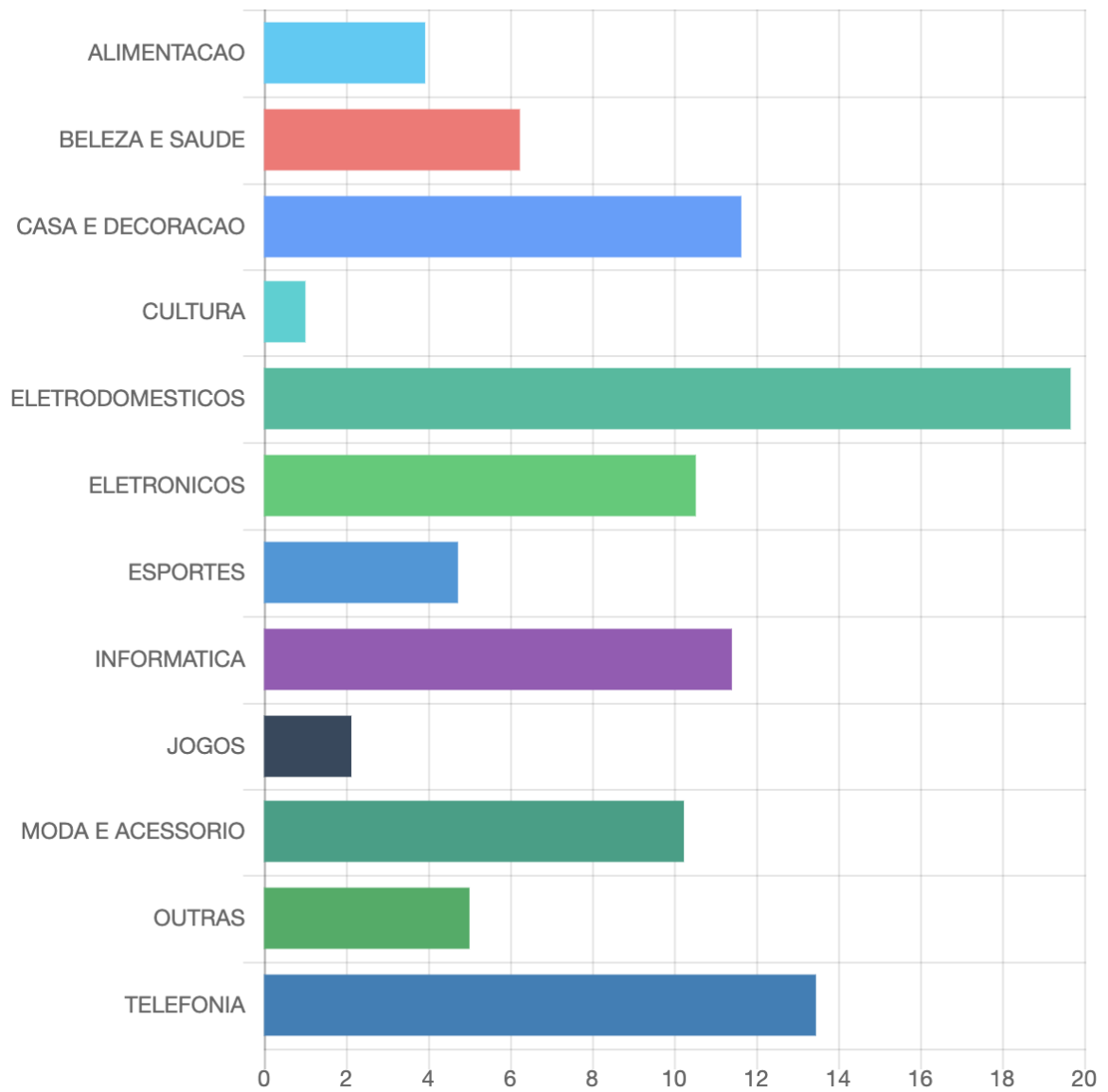
Não gosto muito de fazer compras pelo aplicativo; muitas vezes vejo algo no aplicativo, mas compro pelo notebook para explorar mais. Frequentemente, adiciono itens ao carrinho pelo celular, mas finalizo a compra pelo notebook, onde a visualização é melhor em uma tela grande. Embora o celular torne a navegação e a escolha de produtos mais rápidas, a finalização é mais segura e detalhada no notebook. (Resposta 2).

Costumo olhar produtos no celular quando estou em trânsito, como no ônibus, e uso o notebook principalmente em casa, onde passo pouco tempo. Sinto-me mais seguro usando um notebook conectado a uma rede. (Resposta 3).

Além disso, foi necessário identificar as categorias de produtos mais consumidas pelos entrevistados e compará-las com o mercado em geral.

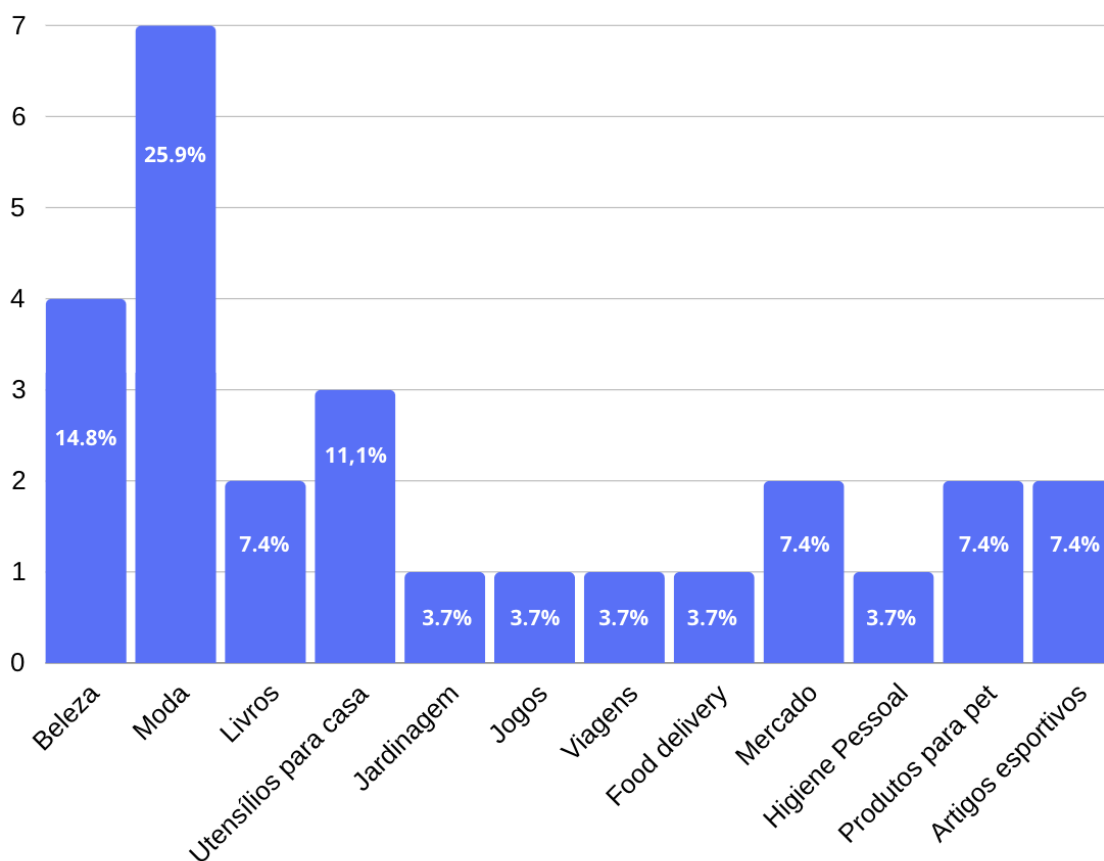
Enquanto os resultados da pesquisa indicaram interesse principalmente nos setores de beleza, moda e utensílios para casa, a Abcomm revela que os setores de maior faturamento são eletrodomésticos, telefonia, casa e decoração. Apesar das diferentes terminologias, acreditei que "casa e decoração" e "utensílios domésticos" estão contextualmente relacionados, por isso os considerei como uma única categoria: utensílios para casa.

Gráfico 5 - Perfil dos Compradores Online (Faturamento) - Categorias de consumo



Fonte: Abcomm (2024).

Gráfico 6 - Distribuição das categorias de consumo dos usuários entrevistados



Fonte: Elaborado pela autora (2024).

Com esses dados, já sabemos que nossos usuários apresentam as seguintes características: preferências de categorias de compras, gênero predominante, faixa etária, região do país e dispositivos utilizados. Essas informações da amostragem estão bem alinhadas com os dados fornecidos pelo mercado, divergindo mais nas categorias de consumo, mas convergindo na categoria de utensílios para casa.

Esses dados são extremamente relevantes para o desenvolvimento do projeto e ajudaram a embasar a taxonomia em diferentes contextos na Fase 2. Com essas informações em mente, vamos agora analisar mais a fundo as informações fornecidas pelos usuários da pesquisa. Vamos pensar na navegação dos sites, o sentimento do usuário, suas dores e o cenário ideal.

Analisando as respostas referentes ao cotidiano do usuário, pontuei de forma resumida as afirmações de cada um:

- Pesquisam antes de comprar.
- Compram mais pela internet devido a cupons, descontos, promoções e preços mais baixos.

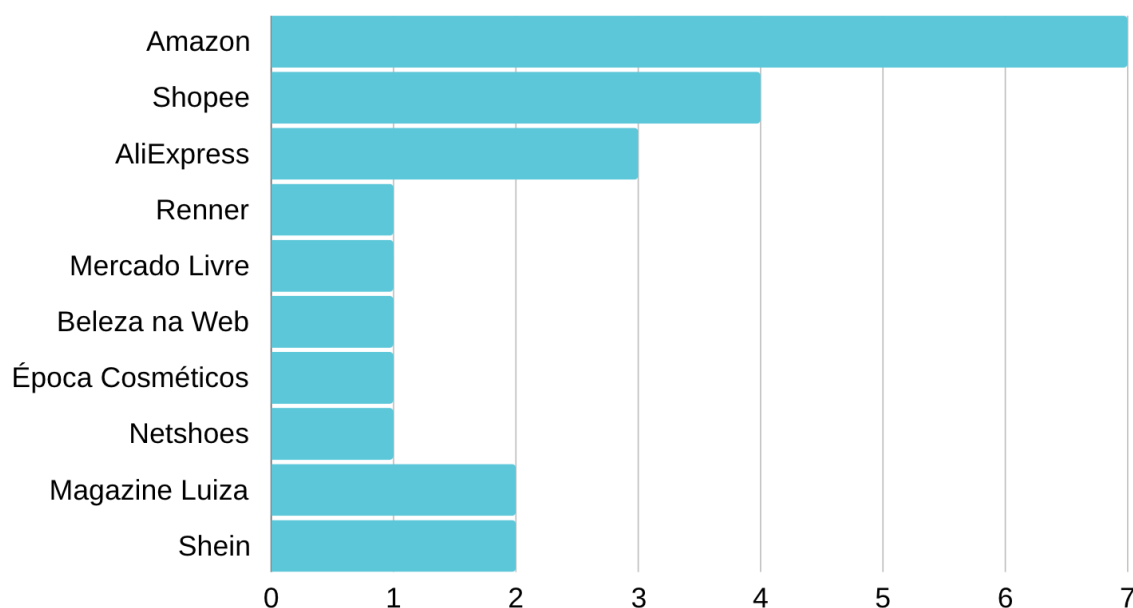
- Itens básicos geralmente compram em lojas físicas, especialmente quando, nas palavras do usuário, “está com preguiça”.
- Preferência em consumir produtos que vê nas redes sociais, com um destaque ao TikTok.
- Só sai de casa para comprar algo se for para o mercado ou para experimentar roupas, e as experimenta para comprar online depois.
- Normalmente, fazem compras online para quase tudo que precisa, exceto produtos de supermercado.
- E um usuário relatou que prefere consumir produtos através de lojas especializadas.

Com essas informações, percebemos que as redes sociais, especialmente o TikTok, devido aos seus vídeos curtos e pontuais, têm um impacto de influência muito forte sobre o usuário. Embora o mercado – aqui, refere-se às mercadorias domésticas da casa - tenha aparecido nos dados com uma relevância de 7,4% na categoria de compras, este ainda é voltado a compras muito pontuais, conforme foi afirmado por um usuário da pesquisa, como azeite ou produtos de limpeza, que esporadicamente aparecem em promoção.

No que diz respeito aos valores, a preferência pelo consumo online se deve a fatores, como: preço, descontos, promoções e cupons. Muitos desses aspectos são difíceis de serem encontrados nas lojas físicas. Mesmo que os preços e descontos sofram variações, no ambiente online, eles são muito mais voláteis e dinâmicos, proporcionando ao usuário mais liberdade para navegar e visitar o mesmo site em busca das melhores ofertas.

Buscando validar se realmente os fatores mencionados anteriormente têm grande impacto nas compras, perguntei aos usuários quais são os sites mais consumidos e o motivo dessa preferência. Veja no gráfico a seguir com as respectivas justificativas dadas pelos usuários:

Gráfico 7 - Sites de preferência para fazer compras online



Fonte: Elaborado pela autora (2024).

Como é possível observar, a gigante Amazon se mostrou ser a opção preferida dos usuários, destacando-se pela entrega rápida, que foi o maior motivador para sua escolha. O fator preço também foi mencionado com frequência. A Shopee, uma empresa chinesa, ficou em segundo lugar, com destaque para cupons, preços e variedade de produtos. A AliExpress, também chinesa, ocupou o terceiro lugar, sendo a variedade de produtos e o preço os principais fatores para a compra. Empatados, estão Magazine Luiza e Shein. A Shein se destacou pelos artigos de beleza, enquanto a Magazine Luiza, nacional, não teve um tópico específico de destaque, mas foi a preferência de dois usuários entrevistados.

Para avançarmos, vamos explorar mais a fundo como os usuários percebem a organização dos sites escolhidos e como eles navegam para encontrar e comprar produtos, até chegarem ao sistema de recomendação. O objetivo principal foi investigar se a complexidade na navegação e na organização de um site levaria os usuários a utilizarem o sistema de recomendação para resolver esse problema.

Após questionar sobre a organização dos sites, prosseguimos com a investigação sobre como os usuários interagem com o sistema de recomendação. Abaixo estão alguns trechos relevantes sobre a percepção da organização dos sites mais utilizados, conforme relatado pelos participantes da pesquisa, elencados de “Resposta 1 a 5” conforme cada relato.

Quanto mais fácil for de navegar, encontrar produtos e fazer tudo com rapidez, é melhor. (Resposta 1).

Os fatores fundamentais para minha decisão de compra incluem a facilidade de navegação e a clareza das informações nos sites. Embora alguns sites sejam difíceis de navegar devido a pop-ups e notificações, no geral, são aceitáveis. Sites como Shopee e Amazon Prime, apesar de populares, têm uma organização confusa e uma busca ineficiente, especialmente para os brasileiros. Em contraste, o Mercado Livre atende melhor o público latino-americano. (Resposta 2).

Seguimos muitos padrões de consumo norte-americanos, onde os sites tendem a ser mais minimalistas, influenciando a usabilidade e organização das plataformas. Sites confiáveis e com entrega rápida são essenciais. Contudo, acho caótica a quantidade de informações que esses sites possuem. Mesmo utilizando recursos de busca e filtro, a filtragem em sites como Amazon e Shopee não é muito eficiente, dificultando encontrar o que procuro. (Resposta 3).

A Amazon e a Shopee são mais fáceis de usar, com a Amazon sendo bem-organizada e agradável de navegar. Outros sites, como Magazine Luiza e Americanas, são confusos e desagradáveis de usar. (Resposta 4).

Vejo o site da Shopee e do AliExpress como bastante caóticos. Porém, isso não me impede de comprar. Utilizo outros recursos, como a busca. Por mais que eu ache a navegação e os filtros caóticos, ainda assim dou um jeito de comprar. Mas, se eu não estiver muito interessada em algo, pode ser que essa desorganização me faça desistir da compra. Agora, se for algo que eu realmente precise, como uma calça jeans para um evento importante, mesmo que o site esteja desorganizado, eu vou encontrar um jeito. Continuo explorando até conseguir o que preciso. (Resposta 5).

A seguir, são apresentados alguns trechos relacionados aos sistemas de recomendação, baseados nas perguntas descritas abaixo. Estas perguntas fazem parte do formulário utilizado na pesquisa, sendo que a justificativa para cada uma delas está disponível no Apêndice D:

Qual a sua opinião sobre as categorias de recomendação dos produtos dos sites que você consome, elas são condizentes?

Sempre são condizentes com as minhas compras e me induz a comprar mais coisas. (Resposta 1).

São similares. As vezes o algoritmo peca. (Resposta 2).

Acho que depende muito do site. Alguns têm essa parte de recomendações bem estruturada, enquanto outros nem tanto. Por exemplo, a Amazon funciona bem para recomendações. Embora eu não goste muito da Amazon em geral, as recomendações são eficientes. Quando você está olhando um livro, eles sugerem outro com um tema parecido ou do mesmo autor, fazendo uma boa correlação com o que pode ser interessante para você comprar além do que está visualizando. (Resposta 3).

Como o sistema de recomendação auxilia você? Ele tem algum peso decisivo na sua escolha? Por quê?

Sim, auxilia na oportunidade de encontrar novos websites e variações de preço+entrega. (Resposta 1).

Peso decisivo, não. O sistema de recomendação não tem um peso decisivo para mim, mas eu o uso para encontrar o que estou procurando. Pode ser tanto um complemento quanto o produto em si. Nos exemplos sobre roupas, acho que a recomendação funciona melhor do que a busca. Então, às vezes, eu o utilizo para encontrar peças específicas. Em outros contextos, uso o sistema de recomendação para encontrar complementos. (Resposta 2).

Eu sou uma pessoa completamente sugestionável. Qualquer campanha de marketing me atrai, especialmente se for de um produto novo que brilha, gira ou faz algo similar a outro produto. Eu acho legal e acabo querendo comprar. Sou facilmente influenciada pelo marketing, como um patinho nas mãos dos marqueteiros. É isso. (Resposta 3).

Normalmente não. Ele auxilia caso eu queira encontrar outros produtos similares e etc, mas não são decisivos na minha escolha final. (Resposta 4).

Me indica produtos parecidos ou que complementam o que eu estou comprando. Se tiver dinheiro para comprar o que me é recomendado, eu acrescento no carrinho. (Resposta 5).

Os trechos selecionados foram comentados pelos usuários, omitindo alguns que, por vezes, desviavam da proposta inicial da pergunta, abordando filtros nos sites e apps e a complexidade na análise de comportamento de compra em cada plataforma. Ao formular esses questionamentos, a intenção era investigar se os usuários navegavam pelas recomendações por serem mais atrativas e alinhadas ao seu padrão de consumo, ou se a falta de organização do site e dificuldades na busca por produtos os levava a utilizar as recomendações como alternativa para encontrar o que desejavam.

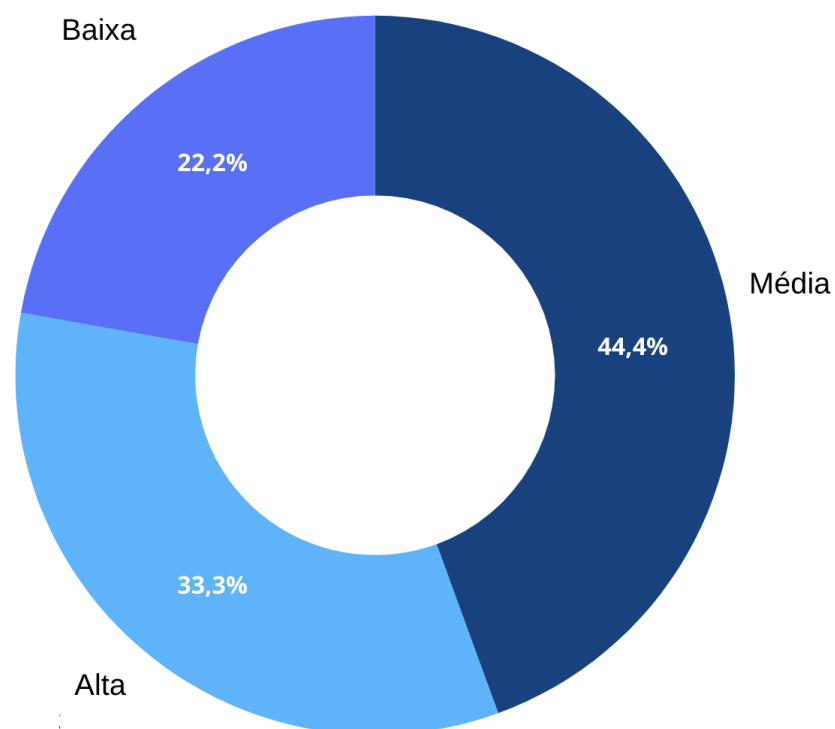
Embora não tenham sido obtidas respostas diretas que validassem uma das hipóteses, foram encontrados indícios de que os usuários utilizavam as recomendações para navegar e localizar os produtos desejados. A persistência dos usuários em encontrar os produtos foi destacada em alguns comentários, indicando que, embora a organização do site seja fundamental, ela não é um impeditivo absoluto na navegação e processo de compra.

É interessante notar que, apesar de parecer senso comum que uma experiência de navegação desorganizada seja muito prejudicial, a pesquisa mostrou que muitos usuários valorizam os benefícios oferecidos, como descontos e cupons, e

continuam insistindo na navegação. O processo de insistência em achar o produto não é imediatamente claro, pois cada usuário pode ter seu próprio caminho dentro da plataforma. No entanto, as informações obtidas ganham força nas respostas sobre o sistema de recomendação. Alguns usuários relataram que são influenciados pelas recomendações e navegam por elas conforme as categorias apresentadas são relevantes aos conteúdos buscados. Um usuário mencionou até que utiliza as recomendações para encontrar roupas, ao invés de usar a busca, que tem se tornado uma dor latente para muitos.

As respostas, portanto, não confirmam que o sistema de recomendação se tornou o caminho principal dos usuários para a tomada de decisão em detrimento da busca, mas destacam que ele é um auxiliar importante no processo de compra, embora não no da decisão final.

Gráfico 8 - Frequência da navegação através das recomendações dos produtos



Fonte: Elaborado pela autora (2024).

Com relação às dores dos usuários, já sabemos que a função de busca nos sites é uma queixa comum, mencionada em diversas respostas que não estavam diretamente relacionadas a essa questão específica. Ao analisar as respostas, que em sua maioria foram curtas, identificamos que a maior dificuldade relatada foi quase unânime: a baixa similaridade entre produtos recomendados. Todos os entrevistados mencionaram este ponto como o principal problema ao utilizar sistemas de recomendação.

Em relação ao cenário ideal, é importante destacar como propósito para as melhorias na experiência do usuário. As respostas para esse cenário foram quase unânimes: os entrevistados desejam um sistema de recomendações com alta precisão de similaridade, além de ter a capacidade de reconhecer o padrão de compra do usuário para sugerir produtos. Além disso, os participantes mencionaram a importância de recomendações do mesmo produto ou de itens similares com descontos, redução de preços e frete grátis, bem como sugestões que compreendam o contexto de compra. Abaixo, destaco alguns trechos que considere relevantes sobre o posicionamento dos usuários.

Seria uma recomendação baseada nos produtos que já comprei. Então, se compro um item, como um action figure, ou produtos para casa, gostaria de ver mais recomendações nessa categoria. Seria útil encontrar produtos similares que estejam com desconto, ofereçam frete grátis ou sejam de um modelo melhor. Por exemplo, se estou comprando um jarro, seria interessante ver recomendações de sementes ou outros produtos relacionados, não necessariamente do mesmo departamento. (Resposta 1).

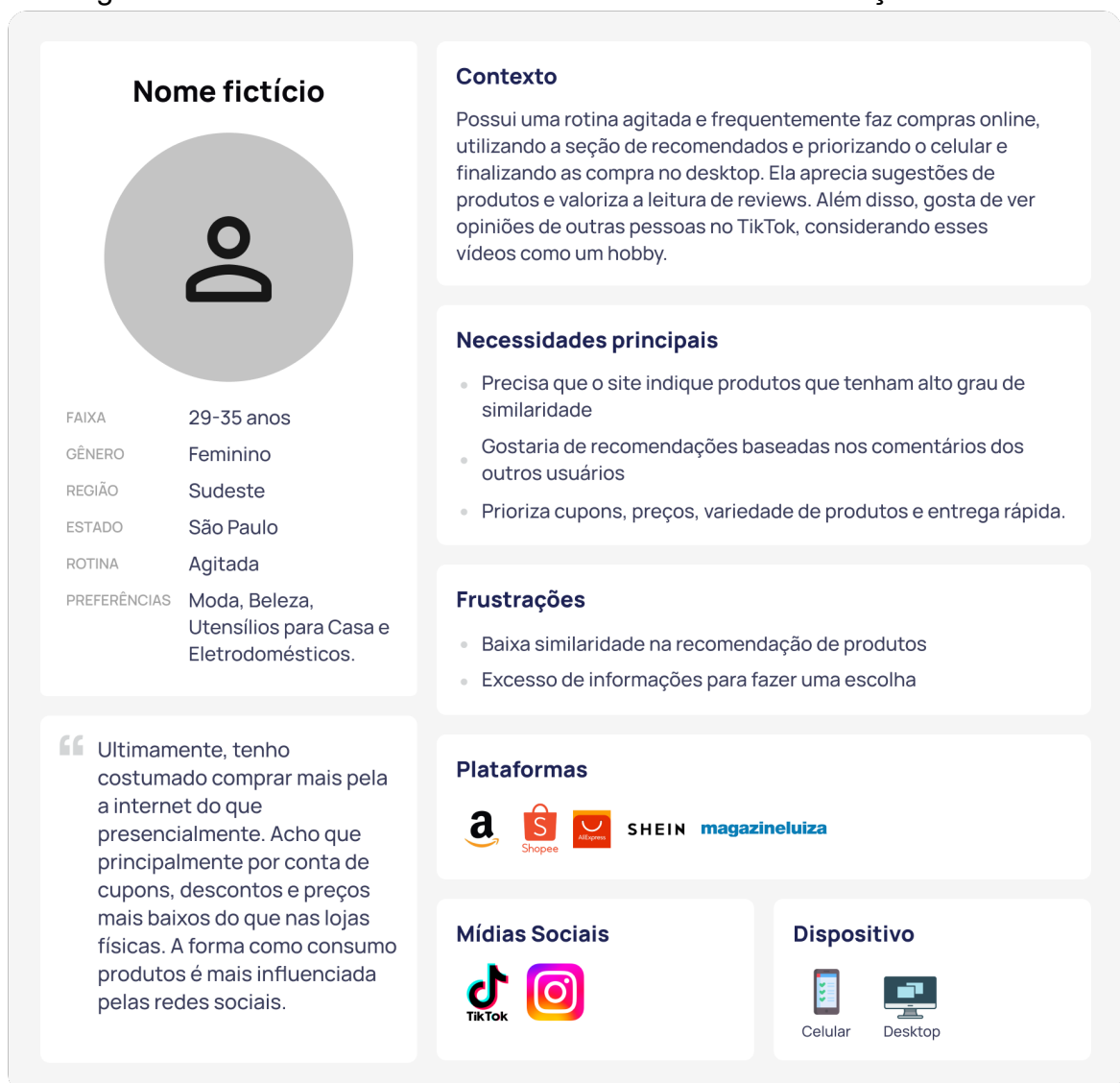
Uma coisa que pensei agora é que talvez você não tenha mencionado o iFood porque é um aplicativo onde as sugestões não pesam muito na escolha. (...) acho que o iFood poderia ser melhor no sentido de mostrar as melhores opções baseadas no que você deseja. Se você está procurando uma pizza margherita, ele poderia destacar as melhores pizzarias que oferecem essa pizza, com base nos comentários e avaliações dos usuários. (Resposta 2).

Ter uma extensão no navegador que sempre haverá de ter a pesquisa de produtos similares em novas plataformas já com as variações de preço, e cálculo de tempo de entrega. (Resposta 3).

É notável que, apesar de não ter incluído trechos específicos aqui para evitar a extensão excessiva do trabalho, as avaliações/comentários, também conhecidas como *reviews*, dos usuários nas plataformas foram frequentemente mencionados como um grande influenciador de compra ao longo da pesquisa. Por essa razão, destaquei o trecho de um usuário que menciona a preferência por recomendações de *food delivery* baseadas nos comentários dos usuários.

Dado todo o insumo, podemos formar um perfil de persona e alinhar alguns pontos com o mercado. A persona, dentro do universo de experiência do usuário, generaliza um grande público, focando no segmento majoritário e trazendo características fortes que ajudam a direcionar o usuário que queremos atingir com melhorias no desenvolvimento de qualquer projeto. Abaixo segue a nossa persona, baseada nos dados retirados da pesquisa e complementada por algumas tendências de mercado.

Figura 2 - UX Persona - Usuário de sistema de recomendação e-commerce



Fonte: Elaborado pela autora (2024).

O perfil da persona é essencial para agrupar e sintetizar informações sobre os usuários, sendo amplamente utilizado no campo da experiência do usuário. Ele auxilia

na definição clara dos objetivos a serem alcançados, centrando-se nas necessidades e frustrações dos usuários. No modelo apresentado, são fornecidas informações anônimas sobre o usuário, embora, em teoria, seja recomendável atribuir um nome fictício, atribuir uma foto e definir uma idade para a persona. A imagem exibida está parcialmente preenchida, com o objetivo de ilustrar a persona desenvolvida para este estudo, mantendo as informações anônimas para preservar a privacidade dos entrevistados e exemplificar como a persona pode ser construída e utilizada.

Temos os dados essenciais para criar as taxonomias, com informações sobre as análises de mercado, as necessidades e frustrações dos usuários, além de uma persona fictícia detalhada. Compreendemos o contexto de uso, incluindo mídias, dispositivos e plataformas, além dos dados demográficos e preferências. Utilizaremos essas informações para aplicar a metodologia proposta para o desenvolvimento de taxonomias, adaptando-as conforme cada contexto específico. Agora, avançaremos para a fase 2 da pesquisa, que se concentrará na estruturação e na governança das taxonomias.

8.2.2 Fase 2: Metodologia para Taxonomias

Vamos definir com mais precisão nossos domínios de trabalho para a taxonomia de conteúdo utilizando a persona. Escolhi três opções: Eletrodomésticos, Utensílios para Casa e Eletroportáteis. A inclusão do primeiro foi baseada nos dados fornecidos pela Abcomm, enquanto os outros são resultados da pesquisa. De todos os tópicos, o único que descartei foi "Moda", pois considerei que sua inclusão tornaria o trabalho excessivamente extenso. Utilizaremos também uma informação crucial sobre a persona para direcionar nosso foco: a frustração relacionada à baixa similaridade na recomendação de produtos. É essa necessidade específica que a taxonomia de conteúdo buscará atender.

- **Recortes de cada categoria:**

Eletrodoméstico: é um vasto universo, e diferentes empresas utilizam termos variados, como "Home Appliances" e "Household" - termo utilizado pela IAB. É importante mencionarmos nomes em inglês, pois, geralmente, essa é a linguagem mais comum nas taxonomias. A maioria das taxonomias feitas e disponíveis para consulta e análise são provenientes dos norte-americanos. Isso também se reflete nos

dados: ao treinar um sistema, é possível enfrentar um cenário onde a língua inglesa será utilizada para atingir seus objetivos.

Para este projeto, vamos usar somente o nome "Eletrodomésticos". Dentro desse contexto, temos uma ampla gama de produtos e aplicações. Considerando que nossa persona é brasileira e vive em um país tropical, escolhi o ar condicionado como o eletrodoméstico a ser trabalhado.

Então, vai ficar assim:

Eletrodoméstico

Ar condicionado

Utensílios para Casa: nesse caso, utensílios para casa em si já é um recorte de algo maior: casa. A pesquisa não deixou claro que tipos de utensílios são esses e para o que servem, contudo, podemos inferir que se trata de organização - essa é uma inferência feita devido às respostas relacionadas ao cotidiano do usuário, em que sua maioria relatou ter uma vida agitada ou relativamente flexível. Olhando para o mercado, a Amazon, a plataforma que a nossa persona mais utiliza divide o universo de Casa, como:

Casa

Ar e Ventilação

Banho

Cama, Mesa e Banho

Cozinha

Decoração para Casa

Ferros, Vaporizadores e Acessórios

Iluminação

Móveis

Quadros e Obras de Arte

Organização e Armazenamento

Produtos de Limpeza

Ao explorar a categoria de Organização e Armazenamento, notamos um universo tão extenso quanto o de Casa, abrangendo opções como Organização para Cozinha, Organização de Garagem, Caixas, Cestos e Baús. Dada essa amplitude, a taxonomia seria enorme e exigiria uma equipe dedicada de taxonomistas para cobrir apenas o segmento de Organização e Armazenamento. Felizmente, empresas como

a Amazon possuem equipes extensivas de taxonomistas para desenvolver essas categorias e treinar seus sistemas de recomendação.

A categorização da Amazon é mais abrangente, mas é útil ter um comparativo para definir o melhor caminho a seguir e estabelecer o universo desejado para esse contexto. Assim, decidi fazer um comparativo com a Shopee, a segunda plataforma mencionada pela persona. Inicialmente, foi difícil encontrar a categoria Casa, possivelmente porque a Shopee destaca apenas as categorias mais buscadas. Contudo, ao utilizar a busca e explorar um pouco a plataforma, ficou claro por que é a segunda opção preferida dos usuários: a variedade e disposição dos produtos são vastas, permitindo até passar horas navegando.

Casa

- Decoração (2mi+)
- Ferramentas e Melhorias para a Casa (1mi+)
- Artigos de Cozinha (997mil+)
- Louça (860mil+)
- Artigos de Festa (768mil+)
- Móveis (401 mil+)
- Iluminação (364mil+)
- Organizadores para Casa (341 mil+)
- Artigos de Cuidados com a Casa (327mil+)
- Roupas de Cama (283mil+)
- Banheiros (276mil+)
- Jardinagem (274mil+)

Analisando ambos os cenários, percebi que, embora usem nomes diferentes, tanto a Amazon quanto a Shopee oferecem categorias semelhantes. Ambos coincidem em trazer "Móveis" e "Iluminação". Decidi focar em Iluminação por ser um tópico interessante, especialmente após explorar o TikTok. Nos vídeos sobre móveis, a abordagem é mais literal, destacando materiais e promoções de lojas. Um ponto importante é que, para promoções, os usuários preferem quando são mostradas pela própria plataforma. Já os vídeos sobre iluminação oferecem dicas de ambientação, cores, economia de energia e marcas, alinhando-se bem mais com a nossa persona, conforme a pesquisa.

Então, vai ficar assim:

Casa

Iluminação

Utensílios para iluminação

Agora temos tudo o que precisamos para iniciar a etapa de Governança.

- **Propósito:** a definição do propósito pode parecer simples, mas é, de fato, o passo mais crucial antes de iniciar a construção de uma taxonomia. Nesta etapa, é fundamental ter clareza sobre a finalidade da taxonomia e como ela será implementada no processo de um sistema. No caso de uma taxonomia de conteúdo, ela é geralmente utilizada para treinamentos de aprendizado de máquina (*machine learning*) e análise de sentimentos. A definição das taxonomias de conteúdo tem como objetivo delinear o funcionamento dentro do aprendizado de máquina para recomendar conteúdo. O propósito das duas taxonomias de conteúdo (Eletrodomésticos e Utensílios para iluminação) é o mesmo, ou seja, treinar para a máquina recomendar conteúdos relacionados.
- **Nomenclatura:** para a área da Biblioteconomia, falar sobre nomenclaturas pode parecer natural. Utilizei esse termo porque, além da Biblioteconomia e Ciência da Informação, poucos conhecem o termo Catalogação. Embora a nomenclatura seja mais simples que o processo de catalogação, ambos têm um viés semelhante, que é rotular uma informação.

Para taxonomias de conteúdo, é importante que as classes tenham nomes simples. Frases extensas devem ser evitadas, a menos que seja um projeto de uso único e essa taxonomia não possa ser utilizada em larga escala. Embora o uso seja delimitado para aprendizado de máquina, gostaria de abordar a análise de sentimentos. A nomenclatura das classes precisa ser neutra o suficiente para "receber" o sentimento da máquina. Por exemplo:

Moda:

EVITAR: "material com baixa qualidade"

"material com pouca durabilidade".

USAR: "Durabilidade".

Isso permite que a análise de sentimentos possa ser avaliada como positiva, negativa ou neutra.

Para a área de Ciência da Informação, isso pode parecer intuitivo e natural. No entanto, no contexto de taxonomias, que transcendem várias áreas e setores, outras áreas podem solicitar nomes que expressam um "sentimento". É importante que o taxonomista negue esse tipo de demanda e justifique, pois isso pode enviesar a máquina com o que chamamos de "*bias*" ou viés.

△ Quando o propósito de uso é para treinamento dentro de uma IA, é imprescindível que a nomenclatura seja a mais neutra possível.

Em resumo, a nomenclatura das 3 taxonomias de conteúdo:

Quadro 4 - nomenclatura das taxonomias de conteúdo

Categoria 1	Categoria 2
Eletrodoméstico Ar condicionado	Casa Iluminação Utensílios para iluminação

Fonte: Elaborado pela autora (2024).

Ao longo do trabalho, você verá toda a estrutura de taxonomia para essas categorias. É importante ressaltar a importância que me preocupei em manter os nomes os mais neutros possíveis em cada uma delas, e alguns até mais diretos e menos detalhados. Com esse nível de objetividade, a máquina pode relacionar melhor os conteúdos, clusterizar e recomendar com mais eficácia.

Definição de tamanho: no caso da taxonomia de conteúdo, quanto mais detalhada e exaustiva, melhor. A máquina precisa dessas nuances e minuciosidades. No entanto, para não sobrecarregar o projeto, e considerando que uma descrição e classificação tão meticulosa exigiria uma equipe dedicada, foi definido que a granularidade máxima será de até 10 níveis. Veja um exemplo abaixo:

Quadro 5 – Definição de tamanho das taxonomias de conteúdo

Níveis	Categoria 1	Categoria 2
--------	-------------	-------------

Nível 1	Eletrodoméstico	Casa
Nível 2	Ar condicionado	Iluminação
Nível 3		Utensílios para iluminação
e assim por diante		

Fonte: Elaborado pela autora (2024).

⚠ **Quanto mais geral for uma classe da taxonomia, menor será seu grau de precisão e maior será o número de informações capturadas.**

⚠ **Quanto mais específica for uma classe da taxonomia, maior será seu grau de precisão e menor será o número de informações capturadas.**

Para a taxonomia de conteúdo, o principal objetivo é treinar o sistema e obter uma volumetria relevante de dados para o treinamento. Quanto maior o número de informações capturadas, melhor. No entanto, é importante reconhecer que alcançar uma precisão alta dessas informações é um grande desafio.

- **Definição de responsabilidade:** nesse contexto, a tomada de decisão e as escolhas são de responsabilidade do taxonomista, que está em contato constante com os dados e o sistema. É crucial que esse profissional não se perca em um foco excessivo na informação e nos dados para a estrutura da taxonomia. Embora o objetivo seja atender ao processo de aprendizado de máquina, é essencial lembrar que o usuário final será quem utilizará os resultados em um sistema de recomendação. A responsabilidade recai sobre o taxonomista, mas é necessário manter sempre em mente o usuário e as necessidades de negócio do projeto.

Agora que estabelecemos a governança, vamos colocar a teoria em prática. É importante lembrar que os exemplos dos dados a seguir são fictícios, baseados em uma base de dados pessoal que possuo há muito tempo. No entanto, eles servirão

perfeitamente para demonstrar como a taxonomia funciona no cenário de aprendizado de máquina.

Vamos começar a construir a taxonomia utilizando as categorias escolhidas previamente: Eletrodomésticos e Utensílios para Casa. Focaremos em níveis de granularidade, buscando atingir um máximo de 10 níveis para manter a complexidade gerenciável.

- **Análise dos Dados:** eletrodomésticos – ar condicionado

Adotei uma abordagem menos detalhada ao descrever as partes do ar condicionado, pois o próprio aparelho é bastante simples. Ele geralmente possui botões, um display e um controle remoto, com um design simples e poucas interações com o usuário. Já uma geladeira, por exemplo, exigiria um detalhamento maior, incluindo gavetas, portas, prateleiras, iluminação, entre outros componentes. Esse seria um processo bem mais extenso.

Criei imagens da árvore taxonômica, mas as letras ficaram pequenas, dificultando a visualização. Por isso, explicitarei a taxonomia da forma a seguir para maior clareza. A estrutura taxonômica, não apenas desta, mas de todas as feitas neste trabalho, está disponível no apêndice.

Eletrodoméstico (nível 1)

- ↳ Ar condicionado (nível 2)
 - ↳ Acessibilidade (nível 3)
 - ↳ Instalação (nível 3)
 - ↳ Performance (nível 3)
 - ↳ Durabilidade (nível 3)
 - ↳ Manutenção (nível 3)
 - ↳ Consumo de energia (nível 3)
 - ↳ Classe A (nível 4)
 - ↳ Classe B (nível 4)
 - ↳ Classe C (nível 4)
 - ↳ Tipo de Ar condicionado (nível 3)
 - ↳ Split (nível 4)
 - ↳ Janela (nível 4)

- ↳ Portátil (nível 4)
- ↳ Conectividade (nível 3)
 - ↳ Wi-fi (nível 4)
 - ↳ App (nível 4)
 - ↳ Assistente de voz (nível 4)
- ↳ Recursos Adicionais (nível 3)
 - ↳ Display (nível 4)
 - ↳ LED (nível 5)
 - ↳ Timer (nível 4)
 - ↳ Modo silencioso (nível 4)
 - ↳ Controle remoto (nível 4)
- ↳ Dimensões (nível 3)
 - ↳ Compacta (nível 4)
 - ↳ Grande (nível 4)
 - ↳ Pequena (nível 4)

Algumas categorias, como acessibilidade, instalação, performance, durabilidade e manutenção, não possuem subgrupos. Isso ocorre porque esses grupos são autossuficientes e, em cenários de treinamento de máquina e/ou análise de sentimentos, podem, sozinhos, abarcar uma quantidade relevante de informações e fornecer os dados necessários. Esses tipos de grupos são muito bons quando a informação é genérica.

Vamos analisar alguns dados para ver na prática:

Planilha 1 - Termos e respectivas frequências para ar condicionado

Termos	Frequência
gela	6134
eficiente	813
potente	582
resfria	545
esfria	503
frio	355
gelar	327

ventilador	197
ventilação	132
evaporadora	128
conforto	107
resfriamento	99
resfriar	83
aquece	82
ar quente	64
aquecimento	48
ventila	48
performance	47
aquecer	33
ice	23
gelo	19
ótimo funcionamento	17
fraco	16
clima de montanha	9
reduzir a umidade	6
produto com bom desempenho	5
função de resfriar	5
esquenta	5
desumidificação	4
falta de umidade	3
potência	1
desemudifica muito o ar	1
supergelado	1
grande fluxo de ar	1
congela	1

Fonte: Elaborado pela autora (2024)

Considerando o contexto de e-commerce e recomendação de produtos, onde os comentários têm um peso significativo na decisão de compra dos consumidores, alguns dos termos nas planilhas são extratos fictícios de comentários online igualmente fictícios sobre produtos, adicionados para proporcionar mais exemplos e volume nas planilhas.

Na planilha acima, temos 36 termos retirados de textos diversos. Esse é um dos resultados do processamento de linguagem natural. Após as etapas de *crawling*, *stemming* e clusterização, o time técnico geralmente fornece uma lista imensa de termos isolados e até frases. Isso representa tudo o que foi capturado no processo de enriquecimento de dados para o time de taxonomia. À direita da tabela, há a coluna

de frequência, que indica quantas vezes o mesmo termo foi repetido nos diversos textos. Só a palavra "gela" é capaz de recuperar 6134 informações relacionadas, por ter sido citada muitas vezes.

Palavras isoladas muitas vezes não revelam muito, mas, quando classificadas dentro do grupo de performance, elas podem fornecer à máquina um contexto útil. Por exemplo, termos como "gela", "esquenta", "ventila", "ventilador", "desumidificação", "clima de montanha", "resfria", "eficiente" e "potente" estão todos relacionados à performance do aparelho. Dessa forma, a máquina poderá associar esses termos ao significado de performance, clusterizando essas palavras. O taxonomista, então, pode definir claramente o que esse grupo representa, abrangendo até mesmo usos mais amplos, como o termo "clima de montanha", que também pode se referir a conforto.

O que é conforto, afinal? O taxonomista deve considerar que, se o usuário se sente confortável em seu ambiente ao usar o ar-condicionado, é de senso comum que o aparelho cumpriu sua função. Se cumpriu sua função, teve uma boa performance. Uma coisa leva à outra, e os termos isolados podem comunicar bastante, especialmente se a frequência de uso for alta. Pensar no contexto do usuário e nos possíveis resultados que levaram ao comentário agrega muito valor ao processo de escolha dos termos e ao resultado final de cada classe dentro da taxonomia.

Mais um exemplo de análise, utilizando o grupo conectividade:

Planilha 2 - Termos e respectivas frequências para conectividade

termos	frequência
app	292
wi-fi	134
alexa	107
google	100
wifi	91
smart	61
thing	30
sensor	19
conectividade	13
smartphone	9
controlar à distância	7
wi fi	7
rede	7
celular	7
smartphones	6

conectar	6
integração com ai	6
controlar remotamente	3
equipamentos inteligentes	3
aplicativo	2
desconecta sozinho	1

Fonte: Elaborado pela autora (2024).

Nesse exemplo de conectividade, temos termos muito importantes para dar ao grupo o contexto necessário. Os principais podem ser considerados "app" e "wi-fi", cada um em seus respectivos grupos com o mesmo nome, enquanto "alexa" e "google" são adicionados ao grupo "assistente de voz". Outros termos como "smart", "thing", "controlar à distância" e "desconecta sozinho" podem ser adicionados ao grupo principal de conectividade.

Embora o termo "desconecta sozinho" tenha apenas uma ocorrência, é extremamente relevante considerá-lo, pois deixa claro que está relacionado ao contexto de conectividade.

Os termos "wi-fi" e "wi fi" se referem à mesma coisa, mas acredite, eles não são a mesma coisa. Um hífen faz toda a diferença, e esses detalhes são cruciais na análise de dados. Uma máquina não percebe como um ser humano que a presença ou ausência de um hífen não muda o significado de "wifi". A máquina precisa ser ensinada sobre isso, pois ela lerá esses termos como duas informações distintas, capturando resultados diferentes. Por exemplo, um termo pode ter uma frequência de 134 e o outro de 7. Incluir ambos no grupo "Wi-fi" é válido e essencial. Estar atento a essas variações na escrita das palavras é muito importante.

Aproveitando que a escrita das palavras se mostrou um desafio válido para sempre manter atenção, vamos entrar mais a fundo na análise de ambiguidade, homônimos e sinônimos.

- **Ambiguidade:** termos isolados também podem ser uma armadilha. Por exemplo, o termo "fraco" pode ser ambíguo. Ele pode se referir à performance, mas também pode descrever o material do aparelho, indicando que algumas partes quebram facilmente, ou pode referir-se ao controle remoto ou ao display. "Fraco" é muito amplo! Se, em vez de um termo isolado, tivéssemos algo como "vento fraco" ou "refrigeração

fraca", haveria uma conversa direta com o grupo de performance. Contudo, "fraco" sozinho pode permear diversos contextos. Assim, ao recuperar informações pelos grupos da taxonomia, o usuário pode obter recomendações sobre, por exemplo, a conectividade com um Wi-Fi fraco.

Gostaria de abordar um tema relacionado à ambiguidade e aos cuidados redobrados com esse aspecto quando buscamos trabalhar com o "Contexto de Uso". Foi exposto um pouco sobre o contexto anteriormente, mas é possível criar classes dentro da taxonomia totalmente voltadas para o contexto de uso, alcançando resultados similares aos de um conceito de ontologia. Como o Processamento de Linguagem Natural (PLN) é muito variável e flexível, lidando com a linguagem, é possível moldar a escolha dos termos para ter um viés ainda mais contextual. Veja um exemplo.

Eletrodoméstico (nível 1)

↳ Ar condicionado (nível 2)

↳ Contexto de uso (nível 3)

↳ Casa de praia (nível 4)

↳ Dias de calor (nível 4)

↳ Ambientes (nível 4)

↳ Uso no quarto (nível 5)

↳ Uso no escritório (nível 5)

↳ Uso na sala (nível 5)

Planilha 3 - Termos e respectivas frequências para contexto de uso – ar condicionado

termos	frequência
uso no quarto	1519
dormir	359
escritório	63
praia	30
morar na praia	14
atende perfeitamente em ambientes médios	17

sala	16
dias de calor	13
dias quentes	10
escritório	7
perfeito mesmo nos dias mais quentes	6
gela muito bem	6
está refrescando a casa nestes dias bem quentes	5
morrer na praia	2

Fonte: Elaborado pela autora.

Veja acima alguns termos. Perceba que "uso no quarto" é o termo mais mencionado, com uma frequência de 1519. Podemos classificar "uso no quarto" no grupo "Uso no quarto", assim como "dormir". No entanto, "dormir" é muito abrangente e não necessariamente se refere a dormir no quarto. Contudo, seguindo uma certa lógica, as pessoas geralmente dormem em seus quartos e ligam o ar para dormir. Assim, ao assimilar esses dois termos, podemos criar um grupo "Uso no quarto" que traga informações relacionadas e contextuais. O mesmo raciocínio pode ser aplicado para "escritório" e "sala", mas não para "praia".

O termo "praia" isolado pode significar muitas coisas, como um lugar, um imóvel em um lugar, ou até mesmo ser usado como figura de linguagem para algo que não deu certo. Podemos ver na planilha que há o termo "morrer na praia". Esse termo pode ter sido utilizado pelo usuário em diversos cenários, como processo de delivery, reembolso, ou instalação, entre outras situações de insatisfação. Em cenários complexos, com mais margem para erros do que acertos, o ideal é não classificar a palavra isolada. Devemos buscar a mesma palavra com o que eu chamo de "qualificador". Um qualificador é aquilo que dá mais sentido à palavra, reduzindo a margem para ambiguidade.

Por exemplo, "casa de praia" é consideravelmente mais específico do que apenas "praia", assim como "morar na praia". Classificando esses termos, o sistema recuperará apenas eles, e não "praia" como um todo, evitando essa problemática abrangente.

- **Sinônimos:** uma máquina pode reconhecer que "praia" é uma praia porque fornecemos insumos através de dados que relacionam várias informações com "praia". Contudo, ela se limita a isso. Se tivermos

termos como "litoral", "beira-mar" ou "orla", a máquina possivelmente não os associará com "praia" porque ainda não aprendeu que esses termos são relacionados, excluindo-os das possibilidades de recomendação.

Digamos que nossa base de dados de contexto de uso seja usada para recomendar ar-condicionado para pessoas que moram na praia. Devido à extensividade da linguagem, muitos usuários podem se referir a locais como "praianos" ao invés de "praia". Nesse caso, o conteúdo se perde e não é recomendado, pois é preciso um aprendizado por reforço para a máquina associar os dois termos e expandir a base. Procurar sinônimos ajuda a máquina a expandir as similaridades e enriquecer a clusterização de dados.

- **Homônimos:** na primeira amostra da base de dados, temos o termo "rede". Esse termo pode, de certa forma, insinuar uma relação com outros termos relacionados à conectividade, como Wi-Fi, especialmente porque alguns usuários escrevem "rede Wi-Fi" ou "rede 3G/4G/5G" para se referir à sua conexão. Contudo, homônimos são palavras muito perigosas e devem ser evitadas por completo se não vierem com um qualificador. Homônimos são palavras com a mesma grafia, mas que podem se referir a coisas completamente distintas.

Em nossa base de dados, temos usuários falando sobre o uso em praia. "Rede" também pode estar relacionada a um objeto de descanso para deitar. Assim, ao recomendar um conteúdo, um usuário interessado em entender melhor a conectividade do aparelho poderia se deparar com recomendações sobre redes para descanso e uso na praia. A experiência do usuário seria péssima, e isso colocaria as informações com um baixíssimo grau de similaridade, até mesmo comprometendo o aprendizado da máquina.

Vamos fazer o mesmo exercício para Utensílios para iluminação. Assim, veremos como é fácil aplicar essa abordagem em diferentes cenários e viabilizar seus resultados. Veja a seguir uma sugestão de taxonomia para esse cenário.

Casa (nível 1)

↳ Iluminação (nível 2)

↳ Utensílios para iluminação (nível 3)

- ↳ Lâmpadas (nível 4)
 - ↳ Led (nível 5)
 - ↳ Fluorescentes (nível 5)
 - ↳ Filamento (nível 5)
 - ↳ Smart (nível 5)
- ↳ Luminárias (nível 4)
 - ↳ de teto (nível 5)
 - ↳ de chão (nível 5)
 - ↳ à pilha (nível 5)
 - ↳ com sensor (nível 5)
 - ↳ USB (nível 5)
- ↳ Adornos festivos (nível 4)
 - ↳ Halloween (nível 5)
 - ↳ Natalinos (nível 5)
 - ↳ Ano novo (nível 5)
- ↳ Estética (nível 4)
 - ↳ Gamer (nível 5)
 - ↳ Boho (nível 5)
 - ↳ Vintage (nível 5)

Vamos analisar alguns dados para ver na prática:

Planilha 4 - Termos e respectivas frequências para utensílios para iluminação

termos	frequência
luz	850
conta de luz	697
led	600
muito agradável	576
alexa	530
smart	348
led televisão	100

luz indireta	84
da uma boa economizada na conta de luz	51
não consigo mudar cor	30
luz de led	25
não possui pilha	11
sunset	10
zen	10
controle remoto do led	5
precisa de 3 pilhas	3
excelente descanso de iluminação	2
smartphone	2
conectado	1
a luz chega ser tão forte que quiser pode iluminar o quarto e sala	1
a luz pisca	1
tem uma grande variedade de combinações de cor de iluminação amei	1
a luz chega ser tão forte que quiser pode iluminar o quarto	1
vela	1

Fonte: Elaborado pela autora.

- **Ambiguidade:** o termo com maior frequência é "luz". De certa forma, é compreensível classificar "luz" quando o assunto é iluminação. Contudo, "luz" é muito amplo e nem se encaixaria bem na taxonomia, pois poderia estar em qualquer classe dentro dela. Além disso, "luz" pode se referir a "conta de luz". Se classificarmos a palavra sozinha, ela recuperará tudo que foi mencionado, incluindo conteúdos sobre conta de luz, o que tornaria a recomendação de conteúdo uma bagunça, com baixíssima similaridade.

Tanto "luz" quanto "conta de luz" podem ter relação com iluminação, mas, no fim, não trazem relevância. Ao contrário de "led", que não traz muita ambiguidade, mas uma gama muito forte de conteúdos relacionados a LEDs, podendo gerar uma boa recomendação de produto. O termo "led" sozinho seria responsável por recuperar toda informação relacionada a lâmpadas LED. Contudo, é sempre bom ajudar um pouco

mais a máquina e classificar "led televisão". Em um contexto em que um usuário procure por LEDs para colocar na televisão, a recomendação de conteúdo será ainda mais eficaz, pois focará na televisão, em vez de recomendar tudo relacionado a LED.

- **Sinônimos:** para analisar sinônimos, escolhi uma palavra estrangeira: "smart". Esse termo é comumente usado para se referir à conectividade de objetos inteligentes. Na planilha, temos "conectado" e "smartphone" como sinônimos de "smart". Ao classificar esses termos, ampliamos o campo de recomendação de conteúdo. No entanto, gostaria de fazer um adicional.

O termo "alexa" não é necessariamente um sinônimo de "smart", mas está relacionado. É importante considerar termos relacionados como um tipo de sinônimo para fornecer mais insumos e "inteligência" à máquina. Por exemplo, quando um usuário procura por lâmpadas LED em um site de e-commerce, o sistema pode recomendar produtos controlados pela Alexa, pois foi um termo previamente classificado e aprendido. Isso fornece mais contexto ao usuário e facilita sua busca por produtos.

Homônimos: "Vela" é um homônimo interessante. Pode ser uma parte de uma lâmpada, mas também é um objeto de cera para iluminar. Contudo, apesar de interessante, o ideal é não classificar a palavra "vela" de forma isolada sem um qualificador, como "vela da lâmpada" ou "lâmpada vela", para deixar claro que está relacionada à lâmpada e não ao objeto de cera. Em um sistema, a máquina aprenderá que "vela" e "lâmpada" estão relacionadas porque uma é componente da outra. No entanto, as consequências no sistema de recomendação podem ser muito ruins, causando cenários em que velas de cera são recomendadas em vez de lâmpadas, resultando em grande confusão para o usuário e o sistema.

Apesar de não aparecerem nos dados, algumas categorias podem ser interessantes. Fiz questão de não incluir dados relacionados às categorias "adornos festivos" e "estética" porque, na criação de taxonomias, é comum fazer um grande estudo de usuário e mercado e os dados não mostrarem que esses termos foram mencionados. Mas isso não é um problema. Todo o processo de coleta de dados e enriquecimento é feito de forma constante, então, mesmo que os dados não apareçam de imediato, há a possibilidade de surgir conteúdo relevante conforme o processo de enriquecimento continue. Assim, cada termo será classificado em sua respectiva

classe dentro da taxonomia, passando a existir e recomendando conteúdos relacionados.

É importante mencionar que, embora eu tenha escolhido manualmente cada um dos termos para demonstrar o processo de criação de taxonomias, esse deve ser um processo de reforço e não um trabalho manual constante. No dia a dia, a base de dados é de bilhões de registros. Geralmente, uma amostra para teste e treinamento da máquina gira em torno de milhões. O taxonomista ajuda manualmente com intervenções pontuais, e a amostragem é reduzida para cerca de dez a vinte mil dados. Isso pode parecer muito, mas na prática é uma boa amostragem para de-paras e reforço humanizado.

Para taxonomias de conteúdo em geral, a metodologia de criação precisa seguir cada um desses passos. Por mais simples que pareçam, não há um consenso mundial sobre as etapas que uma taxonomia deve seguir e os critérios para as principais decisões de seu uso. Como diz Guilherme Pintto (2019), em seu livro "O óbvio precisa ser dito", às vezes, é importante destacar o que parece evidente. Em um cenário de teste e comparação, seria interessante mensurar a qualidade de uma taxonomia construída com e sem esse método específico para reforçar sua governança em um contexto de aprendizado de máquina.

Agora, vamos explorar a **taxonomia de audiência**.

Uma taxonomia de audiência tem como objetivo categorizar diferentes segmentos de público de mercado. É usada principalmente por áreas como marketing e publicidade digital. Seu objetivo é ajudar a personalizar campanhas e organizar essas campanhas de forma a entender o público-alvo de maneira mais profunda. Em geral, essa taxonomia é uma ferramenta estratégica para análise, personalização de publicidade e segmentação de campanhas.

Percebendo esse contexto, é preciso que a linha de raciocínio para a criação dessa taxonomia seja voltada ao consumo de mercado em relação às informações. Ela tem um viés mais mercadológico e, apesar de ser utilizada para segmentar um grupo de usuários, faz isso a partir de seus interesses, comportamentos, hábitos, entre outras categorias relacionadas à audiência. Esse tipo de taxonomia é muito voltado ao mercado em si e a produtos self-service, onde o usuário utiliza plataformas para fazer essa segmentação. Não só empresas utilizam essa taxonomia, mas também usuários que desejam segmentar em plataformas como Pinterest, Meta e Google.

Uma possibilidade de utilização dessa taxonomia é enriquecer o conhecimento sobre o segmento em que o usuário está inserido. Por exemplo, para um usuário ser considerado Gamer, ele precisa preencher certos requisitos definidos por empresas e instituições que se rotulam como Gamer. No nosso exemplo, podemos dizer que usuários nascidos entre 1995 e 2010, residentes de uma região específica como o sudeste do Estado de São Paulo, que consomem jogos como LoL e Valorant, e que têm interesse em notebooks com configurações mais específicas de mercado ou mais elaboradas, seriam considerados Gamers. Esse é um tipo de segmentação feito utilizando uma taxonomia.

Como resultado, o usuário receberá recomendações de produtos e anúncios voltados a notebooks com configurações mais elaboradas e até recomendações de jogos em geral e jogos específicos de interesse. Como uma taxonomia de mercado, geralmente ela é baseada em tendências, estudos e taxonomias mercadológicas, como a taxonomia do IAB. A abordagem é bem diferente justamente porque seu propósito é distinto.

Taxonomias de audiência raramente são usadas para treinamento de Inteligência Artificial porque tendem a atender o público final, ou seja, o usuário que deseja segmentar, enquanto a taxonomia de conteúdo está presente no início e durante o processo. A taxonomia de audiência comumente é um resultado de um processo. Devido às suas características únicas, como neutralidade e assertividade, a taxonomia de conteúdo é mais fácil para treinar IA. Já a taxonomia de audiência trabalha com categorias mais subjetivas, altamente flexíveis e que estão em constante mudança. São taxonomias que estão sempre sendo revisitadas e atualizadas para acompanhar o mercado e as tendências. Lidar com uma taxonomia tão volúvel é arriscado para o treinamento de IA porque os conceitos mudam gradativamente, o que pode comprometer o aprendizado da máquina.

Apesar disso, dentro de um cenário mais amplo, uma taxonomia de audiência fornece insumos suficientes para entender ainda mais o usuário e o mercado, possibilitando até a retroalimentação dentro de um sistema de inteligência artificial. Quando lidamos com uma taxonomia volúvel, é necessário definir um método de criação para garantir que sua estrutura possa acompanhar as mudanças ao longo do tempo. Isso evita a necessidade de criar várias taxonomias com o mesmo propósito. Utilizar uma única árvore taxonômica com uma estrutura base sólida permite abarcar essas mudanças e novos conceitos, mantendo a liberdade de atualização constante.

Para começar, é primordial definir o propósito de uma taxonomia de audiência.

Propósito: é preciso deixar muito claro que o propósito de uma taxonomia de audiência não é para ser utilizada no treinamento de inteligência artificial. Esse é um ponto crucial, pois se essa taxonomia de audiência em algum momento for utilizada com o propósito de treinamento de máquina, é necessário que as mudanças, atualizações de conceitos e afins ocorram com uma frequência bem baixa, quase nula, para que a máquina possa aprender adequadamente. Contudo, raramente se utiliza uma taxonomia de audiência para esse propósito.

A definição do propósito é tratar do assunto que se decide abordar, como informações gerais sobre interesses, hábitos, dados demográficos, aparelhos eletrodomésticos, intenção de compra, entre outros. Esses são os tipos de informações específicas que podem ser exploradas em uma taxonomia de audiência. Geralmente, uma taxonomia de audiência lida com segmentos de mercado que seguem padrões sugeridos pela IAB (Interactive Advertising Bureau). No entanto, nada impede que cada instituição ou empresa desenvolva sua própria estrutura de taxonomia.

Por exemplo, pode-se utilizar a taxonomia da IAB como inspiração ou até mesmo replicá-la, mas também é possível criar suas próprias categorias. Um exemplo disso é o Pinterest, que utiliza uma gama enorme de interesses de usuário baseada no padrão que a empresa encontrou dentro da sua base de dados fornecidos pelos usuários. Ou seja, os interesses encontrados na plataforma Pinterest são ainda mais específicos e voltados ao padrão do que os usuários consomem na plataforma, além dos interesses em geral encontrados no mercado. Isso enriquece a visão de conteúdos a serem segmentados, trazendo mais inspirações e insumos para inovação e captação de novos usuários para a recomendação de anúncios e conteúdo.

Em resumo, a definição do propósito inclui determinar as categorias de mercado que se deseja explorar e elucidar em qual etapa do processo essa taxonomia de audiência será utilizada. No caso deste trabalho, foquei na etapa final do processo, voltada para a segmentação do usuário e entendimento do mercado, visando a recomendação de anúncios e conteúdo.

- **Persona:** uma característica muito interessante da definição da persona no contexto de audiência é que ela não é determinada exclusivamente pelo taxonomista. No contexto deste trabalho, a persona foi baseada em

pesquisas que realizei com usuários que aceitaram ser entrevistados, bem como em dados de mercado que consegui obter.

Geralmente, essa persona vem dos próprios setores de mercado, como publicidade e marketing, que desejam entender melhor um segmento específico. No meu caso, já possuo um segmento de usuários com interesses bastante específicos, sendo um deles o segmento de beleza.

Esse foi o tema que escolhi para a definição do interesse da persona, com o objetivo de direcionar a recomendação dos meus anúncios e conteúdo.

- **Nomenclatura:** os nomes dentro de uma taxonomia de audiência podem ser um pouco mais criativos, contudo, é muito importante que não contenham menções a marcas ou setores específicos. Nos setores de publicidade e marketing, é comum que os usuários, ao contribuírem na criação da taxonomia, queiram inserir marcas e nomes de setores na nomenclatura das classes. No entanto, isso é completamente errado.

Não obstante, uma vantagem de uma taxonomia de audiência é que ela permite mais criatividade e maleabilidade na nomenclatura. Entretanto, é preciso atenção para evitar nomes muito longos ou que contenham apenas uma palavra. Embora esses nomes possam fazer sentido para os usuários que participam da construção da taxonomia, eles podem não ser intuitivos para o usuário final, que não esteve envolvido no processo lógico de categorização. Isso pode atrapalhar a segmentação e a recomendação de conteúdo.

Portanto, é crucial levar em consideração esses aspectos na nomeação das classes da taxonomia. Embora pareça óbvio para a área da Biblioteconomia, esses pontos podem não ser tão evidentes para outras áreas. Por isso, é importante destacar essa questão.

- **Definição de tamanho:** pode parecer contraintuitivo, especialmente para a área da Biblioteconomia, mas a exatidão na construção de classes com um detalhamento excessivo não funciona bem para uma taxonomia de audiência. Essa taxonomia é um ponto de contato direto do usuário final com a informação. Se a informação estiver muito específica em uma classe distante e profunda na hierarquia, o usuário terá dificuldade em encontrar o que deseja. Da mesma forma, se as classes forem muito gerais e superficiais, pode ocorrer a criação de

grupos específicos no mesmo nível que grupos mais gerais, causando estranheza e confusão para o usuário.

Por isso, é crucial definir um tamanho médio para uma taxonomia de audiência, considerando o comportamento do usuário final. É importante perceber até que nível de profundidade o usuário costuma explorar e onde ele tende a parar. A partir dessa observação, pode-se definir um tamanho médio ideal para a taxonomia de audiência, que geralmente varia entre seis a sete níveis.

Neste trabalho, as taxonomias de conteúdo foram apresentadas de forma mais simples e assertiva para exemplificar seu funcionamento, mas elas são, na prática, mais exaustivas e abrangentes, captando uma vasta quantidade de dados para o aprendizado de máquina. Em contraste, as taxonomias de audiência tendem a ser menores, pois precisam ser mais acessíveis e intuitivas para o usuário final. Portanto, sugiro que uma taxonomia de audiência não ultrapasse seis a sete níveis.

- **Definição de responsabilidade:** algo importante a ser mencionado é que a taxonomia de audiência é altamente transitória entre setores e amplamente consumida, não apenas pelo usuário final, mas também por diversos setores. Isso torna fundamental a definição clara de responsabilidades na criação da taxonomia. A responsabilidade principal deve ser do taxonomista, que deve coordenar o processo de construção da taxonomia. Contudo, devido à sua natureza voltada ao mercado e às tendências, é essencial que haja uma cocriação desta taxonomia.

Representantes das áreas de publicidade e marketing devem atuar como cocriadores na conceitualização das classes. Esses profissionais trazem um conhecimento contínuo e atualizado das tendências de mercado e comportamentos do consumidor, o que é vital para a relevância e eficácia da taxonomia de audiência. Dessa forma, a taxonomia pode ser continuamente adaptada e refinada para refletir as mudanças no mercado, garantindo que continue a ser uma ferramenta eficaz para a segmentação e personalização de campanhas.

Agora que foi definido os aspectos iniciais e importantes de uma boa governança de taxonomia de audiência, há materiais suficientes para a criação de uma taxonomia de audiência para o ramo da beleza. Isso permite exemplificar melhor a análise de dados e as questões relacionadas à ambiguidade, sinônimos e

homônimos. Veja abaixo como ficou uma taxonomia de audiência voltada para a persona construída no início deste projeto.

Interesse (nível 1)

↳ Beleza (nível 2)

↳ Autocuidado (nível 3)

↳ Cuidados com a pele (nível 4)

↳ Creme esfoliante (nível 5)

↳ Óleo de limpeza facial (nível 5)

↳ Cuidados com o cabelo (nível 4)

↳ Óleo de cabelo (nível 5)

↳ Maquiagem (nível 4)

↳ Base com retinol (nível 5)

↳ Base com proteção solar (nível 5)

↳ Unhas (nível 3)

↳ Base fortalecedora (nível 5)

Como é perceptível, fiz um recorte ainda mais específico dentro do tema de beleza, focando no universo do autocuidado. Isso é um indicador especialmente relevante porque estamos lidando com o resultado apontado na pesquisa dos usuários e consumo na faixa etária dos 30 anos, uma idade em que há um interesse crescente em questões de autocuidado.

A faixa dos 30 anos é a idade média de nossa persona, e essa taxonomia foi baseada em uma já existente na rede social Pinterest. A imagem a seguir ilustra essa taxonomia.

Figura 3 - Taxonomia de interesse do Pinterest

Escolha interesses relacionados à sua marca para alcançar pessoas através dos Pins com os quais elas interagem.

The screenshot displays a Pinterest interface for selecting interests. On the left, under 'Todos os interesses', there is a search bar and a list of interest categories with checkboxes and counts: 'Beleza (7)', 'Banho e Cuidado Corporal (6)', 'Beleza Masculina (1)', 'Cabelo (5)', 'Cuidados com a Pele (4)', 'Fragrância (1)', and 'Maquiagem (14)'. The 'Maquiagem' category is expanded, showing sub-interests: 'Óleo Essencial', 'Armazenamento de Maquiagem', 'Cópias de Maquiagem', and 'Estilo de Maquiagem (6)'. On the right, a 'Tamanho do público em potencial' (Potential audience size) control is shown, indicating a 'Público mensal dos anúncios ativos' (Active ad monthly audience) of '30 mi+'. A slider is positioned between 'Estreito' (Narrow) and 'Amplio' (Broad). A blue callout box explains that selecting expanded segmentation allows the ad to be shown to other users outside the criteria, improving performance and reaching a larger audience.

Fonte: Pinterest (2024)

- **Análise dos Dados:** na etapa de análise de dados para construir uma taxonomia de audiência, o processo é bastante diferenciado. Os dados já estão bem estruturados em tabelas dentro das bases de dados, e o trabalho é selecionar quais desses dados vão compor a taxonomia. Por exemplo, se estamos trabalhando com uma taxonomia de beleza, buscamos dados relacionados a produtos de cuidados com a pele que estejam literalmente classificados como tal nas tabelas. Esse mesmo princípio se aplica a categorias como limpeza facial, focada em óleos específicos e maquiagem com componentes direcionados ao autocuidado.

Essas informações geralmente estão detalhadas de forma precisa nas tabelas de dados. Ao contrário do processamento de linguagem natural, onde precisamos extrair significado dos dados brutos, aqui lidamos diretamente com dados enriquecidos pelo time especializado. A taxonomia é entregue ao time técnico de enriquecimento de dados, que faz o processo de ligação e enriquecimento dos dados na estrutura taxonômica.

É crucial haver uma comunicação franca entre o taxonomista e o time técnico para definir quais dados serão utilizados em cada categoria. Ao selecionar uma categoria como "beleza > autocuidado > maquiagem" e escolher uma base com componente como o retinol, podemos estimar o número de usuários impactados por essa recomendação de anúncio ou conteúdo. Isso nos ajuda a prospectar o impacto de cada categoria na segmentação de usuários e quantificar a audiência alcançada no mercado.

Apesar do taxonomista não estar diretamente envolvido com o processamento técnico dos dados, é essencial manter uma conversa constante para entender quais dados estão disponíveis na base e garantir que as categorias criadas tenham um número significativo de usuários captados, assegurando assim um impacto substancial no mercado. Exemplo:

Planilha 5 - Tabela de dados para categoria beleza

ID_usuario	brand_name	product_name	product_type
1234567	Revlon	Uniq One	Leave in
891011	L'Oréal	Pro Longer	Shampoo
121314	L'Oréal	Cicatru Renov	Leave in
151617	L'Oréal	Pro Longer	Condicionador
181920	Revlon	Uniq One	Leave in
212223	Revlon	Uniq One Coconut	Leave in
242526	Lancôme	TEINT VISIONNAIRE	Base
272829	Lancôme	TEINT IDOLE	Sérum

Fonte: Elaborado pela autora.

Ao revisar a tabela acima, o taxonomista não necessariamente escolhe os dados para treinamento, como discutimos anteriormente. Em vez disso, tem-se uma ideia de quais dados estão disponíveis e se eles correspondem ao que foi planejado na taxonomia. É comum que alguns dados existam na base, mas não estejam presentes na taxonomia. Por isso, é crucial uma comunicação aberta com o time técnico para fazer ajustes necessários e garantir que os dados disponíveis sejam integrados à taxonomia. Isso ajuda o taxonomista a ter uma compreensão ampla do alcance de informações que a taxonomia de audiência abrange. É importante notar que não é possível ter uma visibilidade completa de todos os usuários prospectados em cada classe até que o produto final esteja pronto.

- **Ambiguidade:** para evitar ambiguidades nas classes de uma taxonomia de audiência, é crucial utilizar qualificadores na nomenclatura das classes. Por exemplo, em vez de usar apenas a palavra "creme", que pode se referir a diferentes tipos, como creme facial, creme hidratante para pele ou creme para rosto, é importante especificar o tipo de creme

dentro da classe. Isso garante que na hora da segmentação não haja confusão sobre o que exatamente está sendo categorizado.

Da mesma forma, foi feito com o termo "óleo", que também necessita de qualificação. Pode ser um óleo de limpeza facial, um óleo massageador ou um óleo para cabelo, cada um direcionado a segmentos específicos de mercado. Mesmo estando dentro de uma classe específica, é essencial qualificar o tipo de óleo para evitar ambiguidades e garantir que a segmentação seja precisa e eficaz.

Portanto, trazer qualificadores claros na nomenclatura das classes é uma prática fundamental para assegurar que cada categoria na taxonomia represente claramente o segmento de mercado que se deseja alcançar, evitando assim qualquer interpretação equivocada durante o processo de segmentação.

- **Sinônimos:** a questão dos sinônimos na taxonomia de audiência muitas vezes passa despercebida, mas é crucial. Muitas análises dessa taxonomia focam na nomenclatura, que não é apenas responsabilidade do taxonomista, mas também de outros setores. É essencial ter uma documentação universal para cada classe da taxonomia de audiência, acompanhada de uma descrição clara. Por exemplo, "Beleza Autocuidado: Unhas - Base Fortalecedora" pode ser descrito como: "Este grupo visa abranger produtos para unhas que fortalecem suas estruturas". Parece simples, mas é fundamental para garantir que a informação seja classificada e representada de forma consistente, evitando a criação de grupos de usuários similares com nomes diferentes. Isso acontece quando diferentes setores tentam atingir o mesmo público-alvo, usando variações na nomenclatura, o que basicamente cria sinônimos na taxonomia de audiência. Portanto, é crucial manter cuidado e clareza na criação e manutenção dessas classes.
- **Homônimos:** homônimos na taxonomia de audiência são semelhantes aos problemas de ambiguidade, pois geralmente estão relacionados à nomenclatura das classes. É essencial sempre incluir um qualificador na criação das nomenclaturas. Por exemplo, "Base Fortalecedora de Unhas" é diferente de uma "Base para Rosto". Mesmo que ambas estejam na categoria de unhas ou maquiagem, usar apenas o termo "base" não seria suficiente, porque cada uma tem propósitos distintos:

uma base para unhas é para fortalecimento, enquanto a outra é mais para finalização, no outro caso uma base é mais leve e outra voltada a total cobertura facial. Portanto, essa qualificação é de suma importância para esclarecer o propósito de cada grupo dentro da taxonomia.

Em geral, a metodologia de criação de taxonomia de audiência visa sempre a segmentação. Um exemplo disso seria a recomendação hipotética de produtos da L'Oréal e Revlon para um usuário específico, cujo perfil foi assimilado na base de dados através do seu histórico de navegação e interesses. Isso ocorre por meio do cruzamento de dados do usuário, seguindo as diretrizes da lei LGPD para proteger suas informações pessoais. As marcas associadas a esses produtos são então recomendadas por meio de anúncios e conteúdos personalizados. Em resumo, esse processo alcança o propósito final de segmentação e recomendação estabelecido por uma taxonomia de audiência.

Em resumo, concluo meu propósito ao definir e criar uma metodologia de taxonomia neste estudo, explorando como eles podem ser aplicados em diferentes cenários e processos. Apesar de não ter sido possível integrá-los diretamente em treinamentos de Inteligência Artificial ou sistemas de prospecção massiva de usuários até o momento, essa abordagem universal na criação de taxonomias promove uma diferença crucial na eficácia dos resultados.

Deixo aberto o caminho para futuros estudos que possam explorar mais profundamente a aplicabilidade dessa metodologia, incluindo experimentações práticas dentro de sistemas específicos. Meu objetivo foi destacar a possibilidade de uma abordagem que pode ser utilizada em larga escala, com diferentes tamanhos e propósitos de taxonomias, visando sempre aumentar sua eficácia em cenários futuros.

É fundamental considerar que a tendência atual do mercado, como indicado pela Statista (2024), mostra que as empresas de comércio eletrônico estão adotando a personalização para compreender melhor as preferências, necessidades e comportamentos individuais dos clientes. Isso inaugura uma era de hiper personalização que se prevê crescer a partir de 2025. Nesse contexto, uma estrutura organizacional bem estabelecida em taxonomia se torna essencial para o gerenciamento eficaz de informações, incluindo aprendizado de máquina e o crescimento contínuo do comércio eletrônico.

9 CONSIDERAÇÕES FINAIS

Este estudo conclui-se com a apreciação do objeto de investigação e das questões que guiaram as análises, atingindo os objetivos inicialmente propostos. O objetivo geral desta pesquisa foi criar uma proposta de uma metodologia para desenvolver taxonomias, tanto para cenários que utilizam processamento de linguagem natural quanto para segmentação e personalização de campanhas e anúncios. Nomeei essas taxonomias como "taxonomia de conteúdo" para informações relacionadas ao processamento de linguagem natural e "taxonomia de audiência" para campanhas e anúncios voltados ao mercado.

O primeiro objetivo específico deste trabalho foi analisar os termos extraídos do processamento de linguagem natural dentro de uma taxonomia, avaliando seu impacto nos resultados de recomendação. Para isso, o segundo objetivo específico foi entender o contexto do usuário, suas necessidades, e como isso orienta a modelagem da informação, desde a captura de dados até a recomendação de produtos, anúncios e conteúdo.

Este estudo visou refletir sobre a importância do usuário como ponto central em todo o processo de personalização online. Em um cenário onde os consumidores esperam que as empresas antecipem suas necessidades, a personalização por si só não será suficiente no futuro. É necessário fornecer soluções hiper personalizadas e experiências complexas para permanecer inovador e atraente em um mercado competitivo.

Estar ciente desse cenário é crucial para se adaptar e reivindicar espaço no mercado de tecnologia, combinando ferramentas analíticas de dados baseadas em inteligência artificial. Isso vai além da segmentação de mercado e recomendação de anúncios e conteúdo, fornecendo ao usuário cenários cada vez mais complexos, analisando seu comportamento e preferências em tempo real. Aproveitando esses dados, é possível criar pontos de contato altamente personalizados, especialmente em relação às recomendações, que são fundamentais na era da personalização e hiper personalização.

O intuito deste trabalho foi trazer desde o conceito de informação e necessidade do usuário até a estrutura de taxonomia, mostrando a importância do processo de organização da informação. A taxonomia é uma estrutura relativamente simples que facilita a categorização, o aprendizado e o consumo de informações

online. Contudo, não encontrei um consenso em minhas pesquisas sobre como as taxonomias devem ser criadas e os melhores métodos de elaboração em cenários de tecnologia, aprendizado de máquina e segmentação de audiência. Sugeri, então, a criação de uma metodologia relativamente simples, mas potente, que pode ser utilizado em larga escala, flexível para diferentes cenários e bastante pontual nas etapas de criação de taxonomia, fazendo toda a diferença em sua aplicabilidade no sistema, especialmente se este for um sistema de recomendação.

Considero que esta pesquisa um ponto de partida para estudos e experimentações sobre taxonomias no cenário de tecnologia e dados, em conjunto com os conhecimentos da Ciência da Informação, como organização da informação, nomenclaturas e questões relacionadas à ambiguidade, homônimos e sinônimos. Isso potencializa, em especial, o cenário de pesquisa em Big Data e aprendizado de máquina.

Deixo aqui uma sugestão para pesquisas futuras, considerando a importância dessas estruturas organizacionais para a experiência do usuário, a eficácia e potencialização dos sistemas e dos resultados, que aqui foram mais focados em recomendação. Espero que a ciência da informação ganhe cada vez mais espaço nas inovações que estão por vir.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE COMÉRCIO ELETRÔNICO. **Principais Indicadores do e-Commerce**. São Paulo: Abcomm, 2023. Disponível em: <https://dados.abcomm.org/numeros-do-ecommerce-brasileiro>. Acesso em: 04 jun. 2024.
- ADAMOPOULOU, E.; MOUSSIADES, L. An overview of chatbot technology. *In: MAGLOGIANNIS, I.; ILIADIS, L.; PIMENIDIS, E. (eds). Artificial Intelligence Applications and Innovations*. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. 2020. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31?ref=blog.min.io#citeas. Acesso em: 27 ago. 2023.
- ALDAYRI, A; ALBATTAH, W. Taxonomy of anomaly detection techniques in crowd scenes. **Sensors**, v. 22, n. 16, p. 6080, 14 ago. 2022. Disponível em: <https://www.mdpi.com/1424-8220/22/16/6080>. Acesso em: 23 jul 2023.
- ALEXOPOULOS, P. **Semantic modeling for data: avoiding pitfalls and breaking dilemmas**. California: O'Reilly Media, 2020. 328 p.
- ALMEIDA, J. Escalabilidade. **Revista Trama Interdisciplinar**, São Paulo, v. 5, n. 3, p. 32-37, dez. 2014. Disponível em: <https://editorarevistas.mackenzie.br/index.php/tint/article/view/7751>. Acesso em: 17 fev. 2024.
- BAPTISTA, S. G.; CUNHA, M. B. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 2, p. 168–184, maio 2007. Disponível em: <https://www.scielo.br/j/pci/a/h6HP4rNKxTby9VZzgzp8qGQ/>. Acesso em: 06 mar. 2024.
- BAILEY, K. D. **Typologies and taxonomies: an introduction to classification techniques**. Thousand Oaks, Calif.: Sage, 1994.
- BARBOSA, O. R. *et al.* Método de taxonomia das paisagens para uso no planejamento ambiental. **Naturae**, v.4, n.1, p.10-15, 2022. Disponível em: <https://www.sapientiae.com.br/index.php/naturae/article/view/184>. Acesso em 20 jun 2024.
- BECHARA, Evanildo. **Moderna gramática portuguesa**. 37. ed. Rio de Janeiro: Nova Fronteira, 2009. 574 p. Edição atualizada pelo novo acordo ortográfico.
- BUGATTI, P. H. **Análise da influência de funções de distância para o processamento de consultas por similaridade em recuperação de imagens por conteúdo**. 2008. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2008. Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-06052008-095020/en.php>. Acesso em: 22 ago 2023.
- BHARADIYA, J. P. Artificial intelligence and the future of web 3.0: opportunities and challenges ahead. **American journal of computer science and technology**, [s.l.],

v. 2, n. 6, p. 91-96, 15 jun. 2023. Disponível em: <https://l1nq.com/DeADW>. Acesso em: 17 fev. 2024.

BRASCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008. São Paulo, **Anais...** São Paulo, USP, 2008. Disponível em: <http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142>. Acesso em: 14 mar. 2024.

CAMBRIA, E.; WHITE, B. **Jumping NLP Curves**: a review of natural language processing research. *IEEE computational intelligence magazine*, maio, 2014. Disponível em: <https://w.sentic.net/jumping-nlp-curves.pdf>. Acesso em: 03 fev 2023.

CAZELLA, S. C.; NUNES, M. A. S. N.; REATEGUI, E. A Ciência da opinião: estado da arte em sistemas de recomendação. *In*: CARVALHO, A. P. L. F.; KOWALTOWSKI, T. (Orgs.). **Jornada de Atualização de Informática-JAI**. Porto Alegre: Sociedade Brasileira de Computação. 2010. p. 161-216. Disponível em: https://scholar.google.com/citations?view_op=view_citation&hl=pt-BR&user=AxhyoRwAAAAJ&citation_for_view=AxhyoRwAAAAJ:UeHWp8X0CEIC. Acesso em: 29 jul 2023.

C3.AI. **O que é um Loop de Feedback?** c2024. Disponível em: <https://c3.ai/glossary/features/feedback-loop/>. Acesso em: 06 jul 2023.

DAVIDSON, J. *et al.* The YouTube video recommendation system. *In*: Proceedings of the fourth ACM conference on recommender systems. **RecSys'10**, New York, NY, USA, p. 293-296, 26 set. 2010. Disponível em: <https://dl.acm.org/doi/abs/10.1145/1864708.1864770>. Acesso em: 03 jul 2023.

DONI, M. V. **Análise de cluster**: métodos hierárquicos e de particionamento. 2004. 93 f. TCC (Graduação) - Curso de Sistemas de Informação, Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie, São Paulo, 2004. Disponível em: <http://meusite.mackenzie.com.br/rogerio/tgi/2004Cluster.PDF>. Acesso em: 07 jul. 2023.

DUBOIS, E.; BLANK, G. **The echo chamber is overstated**: the moderating effect of political interest and diverse media. *Information, Communication & Society*, v. 21, n. 5, p. 729-745, 29 jan. 2018. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1428656>. Acesso em: 05 jul. 2023.

FÉLIX, B. M.; TAVARES, E.; CAVALCANTE, N. W. F. Critical success factors for Big Data adoption in the virtual retail: magazine luiza case study. **Review Of Business Management**, [s.l.], v. 20, n. 1, p. 112-126, jan. 2018. Disponível em: <https://www.scielo.br/j/rbgn/a/YJCMw9MBLV5nVWVT54cPvFp/?lang=en>. Acesso em: 05 jul. 2023.

FRENI, E. J. How AI can help with marketing data standards and taxonomy. **Claravine**, c2024. Disponível em: <https://www.claravine.com/how-ai-can-help-with-marketing-data-standards/>. Acesso em: 14 abr. 2024.

GORINI, M. Classification Vs. Clustering: a practical explanation. **Bismart**, c2024. Disponível em: <https://blog.bismart.com/en/classification-vs.-clustering-a-practical-explanation>. Acesso em 14 ago 2022.

HARDESTY, L. The history of Amazon's recommendation algorithm. **Amazon Science**, 22 nov. 2019. Disponível em: <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm>. Acesso em 14 ago 2022.

JONES, D. Google's Natural Language Processing API Tool for SEO. **Majestic Blog**, 5 mar. 2019. Disponível em: <https://blog.majestic.com/company/googles-natural-language-processing-api-tool-for-seo/>. Acesso em: 02 jan. 2024.

KEINONEN, T. **User-centered design and fundamental need**. Proceedings of the 5th Nordic conference on Human-computer interaction building bridges - NordiCHI '08, 2008. Disponível em: <https://dl.acm.org/doi/abs/10.1145/1463160.1463183>. Acesso em: 10 maio 2023.

KAYID, A. **The role of Artificial Intelligence in future technology**. Department of Computer Science. The German University in Cairo, p. 1-4, 15 mar. 2020. Disponível em: <https://www.researchgate.net/publication/342106972>. Acesso em: 17 set. 2023.

KIELGAST, S.; HUBBARD, B. A. Valor agregado à informação: da teoria à prática. **Revista Ciência da Informação**, v. 26, n. 3, 1997. Disponível em: <https://revista.ibict.br/ciinf/article/view/768>. Acesso em: 10 ago. 2023.

KLEIN, J. T. Typologies of Interdisciplinarity: the boundary work of definition. *In*: FRODEMAN, R. (ed.). **The Oxford handbook of interdisciplinarity**. 2. ed. Oxford Handbooks: Oxford Academic, 6 Mar. 2017. Disponível em: <https://doi.org/10.1093/oxfordhb/9780198733522.013.3>. Acesso em: 10 ago. 2023.

KNIJFF, J.; FRASINCAR, F.; HOGENBOOM, F. Domain taxonomy learning from text: the subsumption method versus hierarchical clustering. **Data & Knowledge Engineering**, [s.l.], v. 83, p. 54-69, jan. 2013. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169023X12000973>. Acesso em: 20 jun. 2024.

LAM, H.; FUNG, G.; LEE, W. A novel method to construct taxonomy electrical appliances based on load signaturesof. **IEEE Transactions On Consumer Electronics**, [s.l.], v. 53, n. 2, p. 653-660, 2007. Disponível em: <https://ieeexplore.ieee.org/abstract/document/4266955>. Acesso em: 20 jun. 2024.

LEE, K. C. **Machine Learning 101: classification vs. clustering**. Classification vs. Clustering. 2020. Disponível em: <https://kevin-c-lee26.medium.com/machine-learning-101-classification-vs-clustering-e11b12c71243>. Acesso em: 10 abril 2023.

LÉVY, P. **O que é o virtual?** 2. ed. Tradução de Paulo Neves. São Paulo: Editora 34, 2011.

LI, Y.; WU, H. A clustering method based on K-Means algorithm. **Physics Procedia**, [s.l.], v. 25, p. 1104-1109, 2012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1875389212006220>. Acesso em: 02 jul. 2023.

LIMA, J. L. O.; ALVARES, L. Organização e representação da informação e do conhecimento. In: ALVARES, L. (org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4 Editores, 2012. p. 21-34. Disponível em: https://www.researchgate.net/publication/281969932_Organizacao_e_representacao_da_informacao_e_do_conhecimento. Acesso em: 02 jul. 2023.

LÜDERS, M. Pushing music: people's continued will to archive versus spotify's will to make them explore. **European Journal of Cultural Studies**, [s.l.], v. 24, n. 4, p. 952-969, 28 jul. 2019. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1367549419862943>. Acesso em: 05 jul. 2023.

MARCO GONZALEZ, V. L. S. L **Recuperação de informação e processamento da linguagem natural**. Porto Alegre: PUCRS - Faculdade de Informática, 2003. Disponível em: <https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/mri-06---gonzales-e-lima-2003.pdf>. Acesso em: 03 jun. 2023.

MEIRA, L. Mudanças e tendências de comportamento do consumidor na era digital. **Surfe Digital**, 17 set. 2019. Disponível em: <https://www.surfedigital.io/blog/comportamento-do-consumidor>. Acesso em: 04 fev. 2019.

MENDES, I.; BENTES PINTO, V. Taxonomia nas áreas da Biblioteconomia e da Ciência da Informação: uma revisão sistemática. **Páginas A&B: arquivos e bibliotecas**, Porto, p. 36–47, 2019. Universidade do Porto, Faculdade de Letras. Disponível em: <https://ojs.letras.up.pt/index.php/paginasaeb/article/view/6379>. Acesso em: 30 jul. 2024.

NASCIMENTO, T. G.; QUINTÃO, P. L. Ferramentas da web 2.0 para a gestão do conhecimento em um ambiente organizacional. **Revista Eletrônica da Faculdade Metodista Granbery**, Juiz de Fora, n. 10, p. 1-26, jan./jun. 2011. Disponível em: <http://re.granbery.edu.br/artigos/NDM1.pdf>. Acesso em: 20 jun. 2022.

NAZARI, Z. *et al.* **A new hierarchical clustering algorithm**. International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, Japan, 2015, p. 148-152, 2015. Disponível em: <https://ieeexplore.ieee.org/abstract/document/7439517>. Acesso em: 20 jun. 2023.

NEIVA, A. C. E-commerce no Brasil: dados de um mercado em expansão. **Edrone**, 16 maio 2024. Disponível em: <https://edrone.me/pt/blog/dados-ecommerce-brasil>. Acesso em: 04 jun. 2024.

NICKERSON, R. C; VARSHNEY, U.; MUNTERMANN, J. A method for taxonomy development and its application in information systems. **European Journal Of Information Systems**, [s.l.], v. 22, n. 3, p. 336-359, maio 2013. Disponível em: <https://www.tandfonline.com/doi/full/10.1057/ejis.2012.26?needAccess=true>. Acesso em: 29 jun. 2024.

NORMAN, D. **The design of everyday things**. Revised and expanded edition. Basic Books: 2013.

O'REILLY, T. What Is Web 2.0? **O'Reilly Media**, 30 set. 2005. Disponível em: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>. Acesso em: 29 jun. 2022.

PAL, K. Spotify: machine learning to personalize the user experience. **Analytics Vidhya**, 18 maio 2021. Disponível em: <https://www.analyticsvidhya.com/blog/2021/05/spotify-machine-learning-to-personalize-the-user-experience/>. Acesso em: 10 maio 2023.

PINTEREST. **Expande o teu negócio com os anúncios do Pinterest**. c2024. Disponível em: https://ads.pinterest.com/advertiser/549766059701/ads/campaign_mode/. Acesso em: 9 ju. 2024.

PINTEREST ENGINEERING. Interest Taxonomy: a knowledge graph management system for content understanding at pinterest. **Medium**, 10 jan. 2020. Disponível em: <https://medium.com/pinterest-engineering/interest-taxonomy-a-knowledge-graph-management-system-for-content-understanding-at-pinterest-a6ae75c203fd>. Acesso em: 01 jul. 2023.

PINTTO, G. **O óbvio também precisa ser dito**. [s./]: Editora Outro Planeta, 2019.

RANGANATHAN, S. R. **As Cinco Leis da Biblioteconomia**. Brasília: Briquet de Lemos, 2009.

RAMESH, A. *et al.* Audience segment expansion using distributed in-database k-means clustering. **International workshop on data mining for online advertising**, 11 ago. 2013. Disponível em: <https://encurtador.com.br/A5zrm>. Acesso em: 02 jul. 2023.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**, v. 1, n. 1, p. 96-105, 2014. Disponível em: https://www.brapci.inf.br/_repositorio/2015/12/pdf_616d318e35_0000015108.pdf. Acesso em: 14 ago. 2023.

ROCHA, J. A. P.; PAULA, C. P. A.; DUARTE, A. B. S. A cognição distribuída como referencial teórico para os estudos de usuários da informação. **Inf. & Soc.**, João Pessoa, v. 26, n. 2, p. 91-105, set. 2016. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/28563>. Acesso em: 06 mar. 2024.

SALES, O. M. M. S; FARIAS, G. B. de; PINTO, V. B; SOUSA, M. R. F de. Arquitetura da informação: estudo e análise da base de dados Public Medical (PubMed), **Biblios**, n. 63, p. 1-12, 2016. Disponível em: <https://www.redalyc.org/journal/161/16146347001/html/>. Acesso em: 10 maio 2023.

SCHAFER, J. B.; KONSTAN, J.; RIEDL, J. **Recommender systems in e-commerce**. Proceedings Of The 1St Acm Conference On Electronic Commerce, [s./], p. 158-166, nov. 1999. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/336992.337035>. Acesso em: 04 jul. 2023.

SCHREPP, M.; HINDERKS, A.; THOMASCHEWSKI, J. Construction of a Benchmark for the User Experience Questionnaire (UEQ). **International Journal Of Interactive Multimedia And Artificial Intelligence**, [S.L.], v. 4, n. 4, p. 40, 2017. Disponível em: <https://reunir.unir.net/handle/123456789/11754>. Acesso em: 06 jun. 2023.

SHARMA, G.; LIJUAN, W. The effects of online service quality of e-commerce Websites on user satisfaction. **The Electronic Library**, [s.l.], v. 33, n. 3, p. 468-485, 1 jun. 2015. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/el-10-2013-0193/full/html>. Acesso em: 10 maio 2023.

SOUTHERN, M. G. TikTok gains traction as a search engine among Gen Z [STUDY]. **Search Engine Journal**, 15 jan. 2024. Disponível em: <https://www.searchenginejournal.com/tiktok-gains-traction-as-a-search-engine-among-gen-z-study/505633/>. Acesso em: 04 mar. 2024.

SOUZA, O.; FERNANDES, J. D. P. B. Discutindo a acessibilidade informacional no ecossistema digital. In: FARIAS, M. G. G.; PINTO, V. B. (orgs.). **Ciência da Informação em contextos**. Fortaleza: Imprensa Universitária, 2020. p. 141-163.

SILVEIRA, L. *et al.* Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 26, p. 1-27, 7 jun. 2021. Disponível em: <https://www.redalyc.org/journal/147/14768130019/14768130019.pdf>. Acesso em: 29 jun. 2024.

STATISTA. **AI Trends: roadmap to 2025**. Roadmap to 2025. c2024. Disponível em: <https://www.statista.com/page/ai-trends-2025-on>. Acesso em: 04 jun. 2024.

STRASSER, A. Delphi method variants in information systems research: taxonomy development and application. **Electronic Journal of Business Research Methods**, v. 15, n. 2, p. 120-133, 1 out. 2017. Disponível em: <https://academic-publishing.org/index.php/ejbrm/article/view/1362>. Acesso em 14 jun 2024.

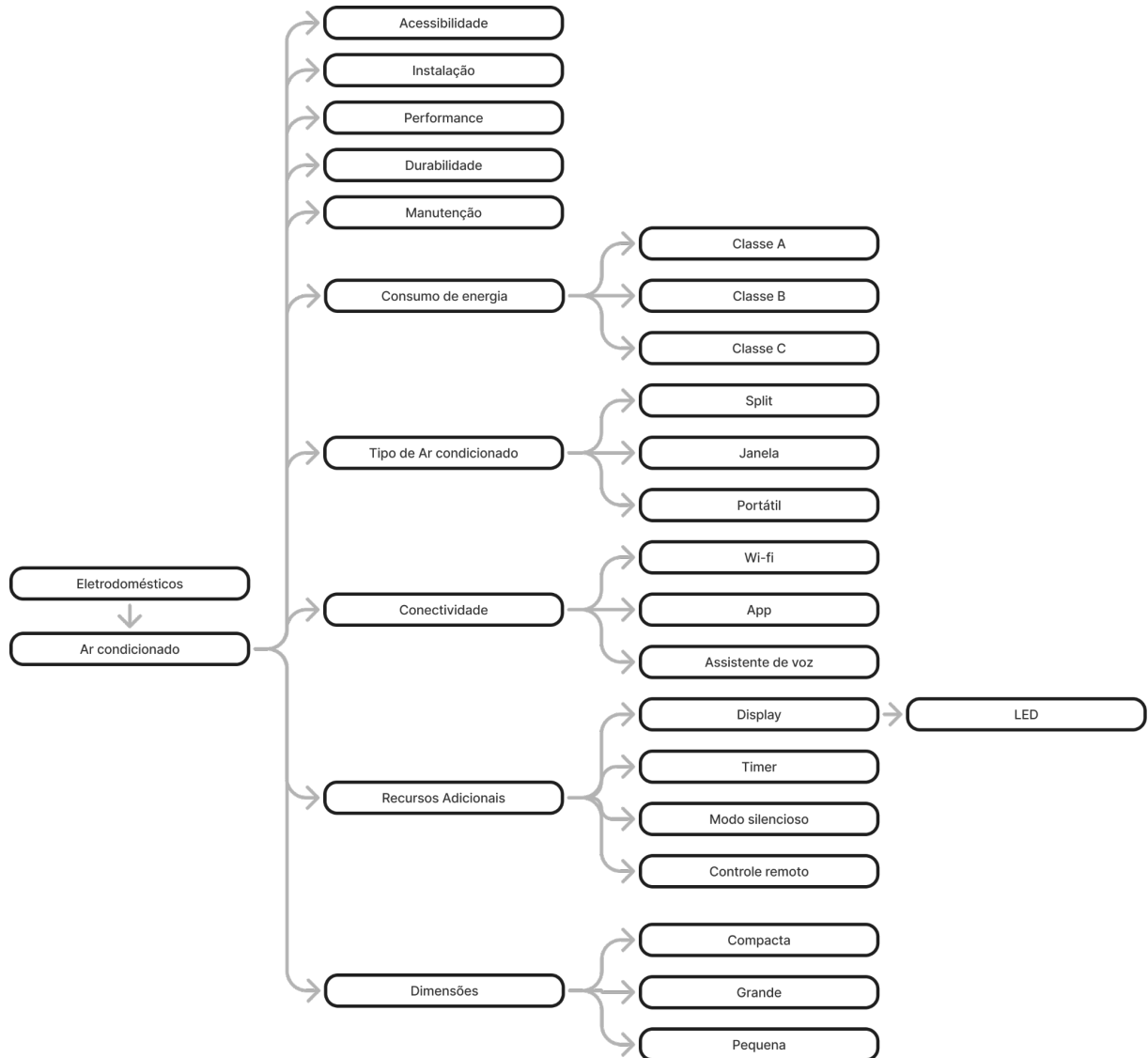
TILIA, C. Como o TikTok virou o novo Google para a Geração Z. **Forbes Tech**, 12 abr. 2024. Disponível em: <https://forbes.com.br/forbes-tech/2024/04/como-o-tiktok- virou-o-novo-google-para-a-geracao-z/>. Acesso em: 19 maio 2024.

USMAN, M. *et al.* Taxonomies in software engineering: a systematic mapping study and a revised taxonomy development method. **Information and software technology**, v. 85, p. 43–59, 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950584917300472>. Acesso em 14 jun 2024

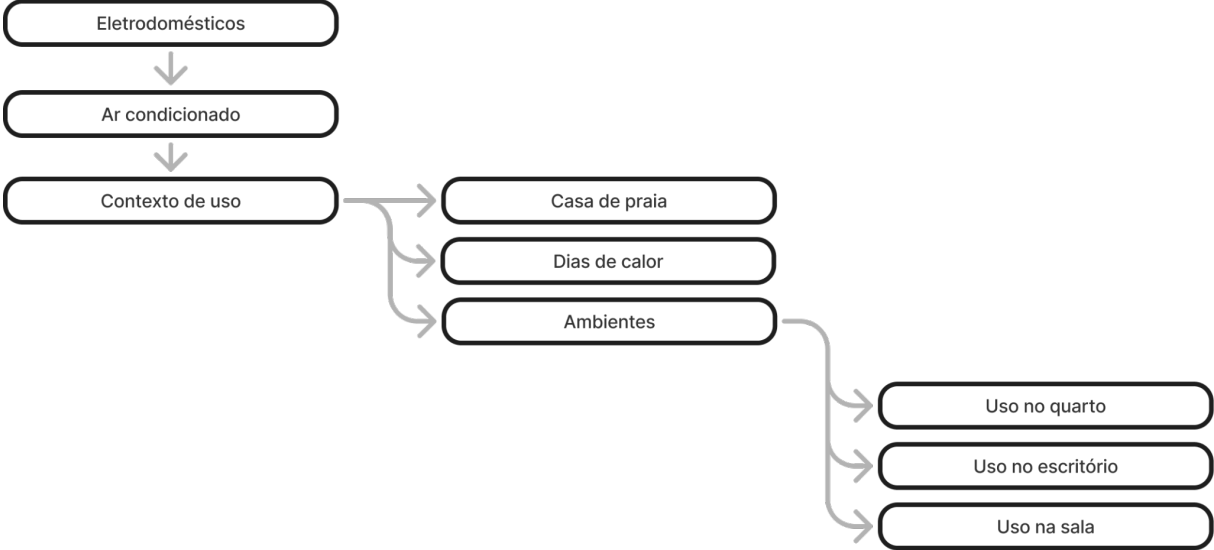
WAL, T. V. Wikipedia folksonomy is a mess with collaborative misunderstanding. **Vanderwal.net**, 29 nov. 2007. Disponível em: <https://www.vanderwal.net/random/entrysel.php?blog=1949>>. Acesso em: 07 mar. 2023.

WELLS, J. Best Practices for Taxonomy Governance. **KM World**, 5 nov. 2018. Disponível em: <https://www.kmworld.com/Articles/News/News/Best-Practices-for-Taxonomy-Governance-128364.aspx>. Acesso em: 14 ago. 2022.

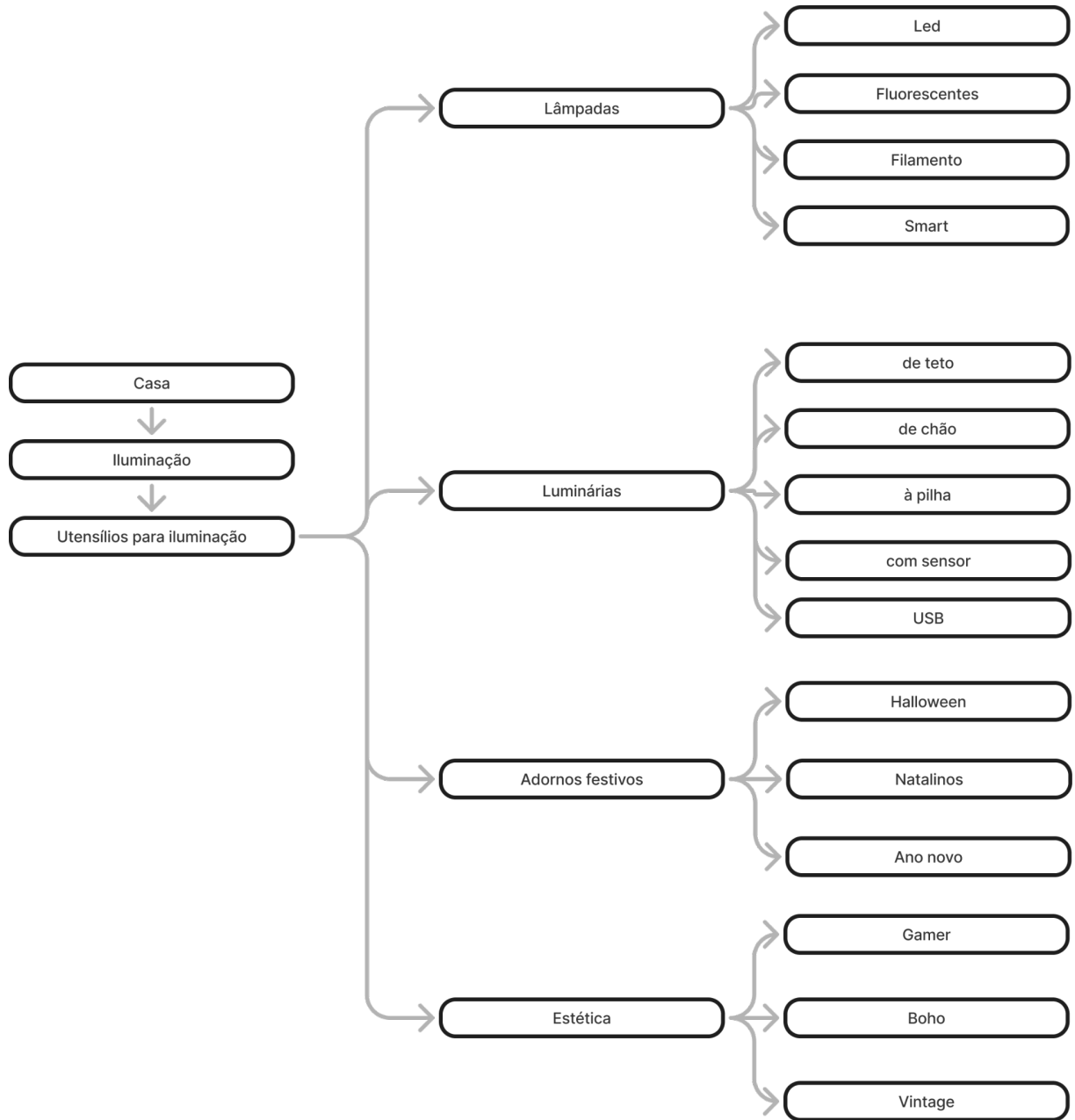
APÊNDICE A - TAXONOMIA DE CONTEÚDO PARA ELETRODOMÉSTICOS – AR-CONDICIONADO



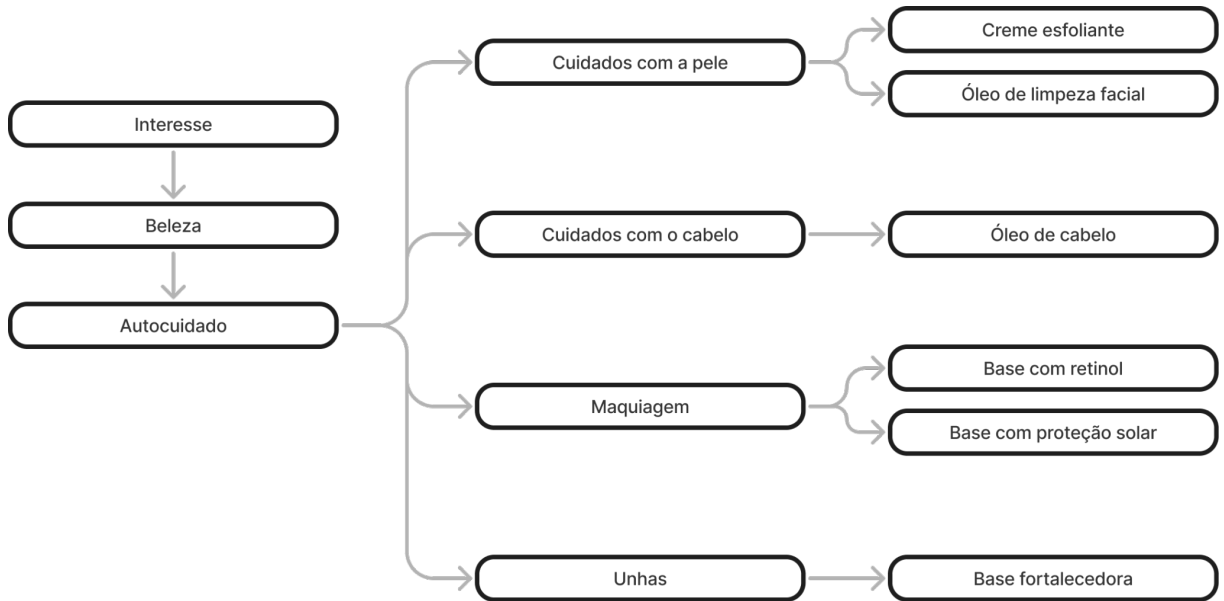
APÊNDICE B - TAXONOMIA DE CONTEÚDO - CONTEXTO DE USO EM AR-CONDICIONADO



APÊNDICE C - TAXONOMIA DE CONTEÚDO - UTENSÍLIOS PARA ILUMINAÇÃO



APÊNDICE D - TAXONOMIA DE AUDIÊNCIA – BELEZA



APÊNDICE D – ROTEIRO DA ENTREVISTA COM OS USUÁRIOS

1. Dados demográficos

- a. Qual o seu nome e idade?
- b. Qual a região do Brasil que você reside atualmente?
- c. Se sentir-se confortável, poderia me informar qual gênero você se identifica?

Justificativa: A seção de dados demográficos permite entender melhor o cenário que o entrevistado está inserido e compreender melhor a amostragem

2. Cotidiano do usuário

- a. Poderia explicar como é seu cotidiano e como prefere consumir produtos? Por exemplo: *Tenho cotidiano agitado e prefiro consumir produtos que vejo pelo instagram.*
- b. Quais os tipos de produtos você mais consome? Poderia categorizá-los? Por exemplo: *produtos de beleza, roupas, acessórios de esporte, suplementos e etc.*
- c. Você costuma comprar produtos utilizando qual dispositivo? Celular? Tablet? Notebook?
- d. Você costuma ter preferência em fazer compras *online* em algum *site*? Se sim, poderia me citar qual(is)?
- e. Qual a motivação na escolha desse(s) sites? Exemplo: *valor, navegação, produtos.*

Justificativa: A seção de cotidiano permite entender o usuário e o contexto de uso dos produtos e plataformas.

3. Arquitetura da Informação

- a. Qual a sua opinião sobre a organização das categorias dos sites que você compra?
- b. Você acha que a organização do site é um fator primordial para consumir produtos nele?
- c. Qual a sua opinião sobre as categorias de recomendação dos produtos dos sites que você consome, elas são condizentes? Exemplo: *Se você está buscando algo sobre livros e a recomendação te sugere uma panela, então as categorias estão destoantes.*

Justificativa: A seção de arquitetura da informação permite entender a navegação do usuário e como é a organização dos produtos na visão dele.

4. Seção de recomendados

- a. Com qual frequência você navega através das recomendações dos produtos que você vê? Pode utilizar numa escala de 0 a 10, ou se preferir a escala de baixa, média e alta frequência.
- b. Como o sistema de recomendação auxilia você? Ele tem algum peso decisivo na sua escolha? Se sim, por quê?
- c. Poderia citar os problemas que você enfrenta ao utilizar a seção de recomendados?

<desdobramento de perguntas em possíveis cenários negativos de utilização dos recomendados>

- Qual sua maior frustração se há pouca similaridade entre as ofertas?
 - Qual sua maior frustração se a seção de recomendados é má posicionada?
 - Qual sua maior frustração se a ordenação das ofertas é difícil e/ou confusa?
- d. Qual seria para você o cenário ideal de recomendação de produtos?

Justificativa: A seção de recomendados permite ter um entendimento geral da visão do usuário em relação aos itens de recomendação das plataformas que consome.

5. Finalização

- a. Há algo que eu não abordei sobre o assunto tratado aqui que você gostaria de aproveitar para expor?
- b. E o nosso encontro, como foi para você?