

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Luzane Ruscher Souto

**MODELO DE REPRESENTAÇÃO DE PRODUTOS DO VAREJO NO COMÉRCIO
ELETRÔNICO:**

uma abordagem a partir da perspectiva da organização e representação do
conhecimento e da informação

Porto Alegre

2024

Luzane Ruscher Souto

**MODELO DE REPRESENTAÇÃO DE PRODUTOS DO VAREJO NO COMÉRCIO
ELETRÔNICO:**

uma abordagem a partir da perspectiva da organização e representação do
conhecimento e da informação

Dissertação de mestrado apresentada como requisito para aprovação em banca examinadora de defesa no Programa de Pós-graduação em Ciência da Informação (PPGCIN) da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, na Linha de Pesquisa Produção e Organização da Informação.

Orientadora: Profa. Dra. Rita do Carmo Ferreira Laipelt

Porto Alegre

2024

CIP - Catalogação na Publicação

Ruscher Souto, Luzane

Modelo de representação de produtos do varejo no comércio eletrônico: uma abordagem a partir da perspectiva da organização e representação do conhecimento e da informação / Luzane Ruscher Souto.

-- 2024.
226 f.

Orientadora: Rita do Carmo Ferreira Laipelt.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Ciência da Informação, Porto Alegre, BR-RS, 2024.

1. Organização e Representação da Informação e do Conhecimento . 2. Sistemas de Organização do Conhecimento (SOC). 3. Análise de Domínio . 4. Comércio eletrônico . 5. Varejo. I. Ferreira Laipelt, Rita do Carmo, orient. II. Título.

Luzane Ruscher Souto

**MODELO DE REPRESENTAÇÃO DE PRODUTOS DO VAREJO NO COMÉRCIO
ELETRÔNICO:**

uma abordagem a partir da perspectiva da organização e representação do
conhecimento e da informação

Dissertação de mestrado apresentada
como requisito para aprovação em banca
examinadora de defesa no Programa de
Pós-graduação em Ciência da Informação
(PPGCIN) da Faculdade de
Biblioteconomia e Comunicação da
Universidade Federal do Rio Grande do
Sul, na Linha de Pesquisa Produção e
Organização da Informação.

Orientadora: Profa. Dra. Rita do Carmo
Ferreira Laipelt

Aprovada em: Porto Alegre, 12 de julho de 2024.

BANCA EXAMINADORA

Prof^a. Dr^a. Rita do Carmo Ferreira Laipelt (UFRGS)

Prof^a. Dr^a. Caterina Marta Groposo Pavão (UFRGS)

Prof. Dr. Thiago Henrique Bragato Barros (UFRGS)

Prof^a. Dr^a. Mariângela Spotti Lopes Fujita (UNESP)

Prof. Dr. Rafael Port da Rocha (suplente) (UFRGS)

AGREDECIMENTOS

Agradeço à minha orientadora por me guiar em decisões importantes, como a mudança do objeto de pesquisa para algo mais relacionado à minha atuação profissional, onde eu já tinha mais experiência. Sou grata pela orientação, compreensão e pelos ensinamentos ao longo do caminho.

À banca examinadora, agradeço pela análise criteriosa e pelas excelentes considerações tanto na qualificação quanto na defesa. Suas observações me ajudaram a melhorar o trabalho, validaram o conhecimento que construí e me deram mais confiança para continuar na carreira acadêmica.

À Universidade Federal do Rio Grande do Sul, onde cursei minha formação de nível superior. A universidade não apenas viabilizou minha educação de forma gratuita, mas também me proporcionou uma experiência acadêmica de alta qualidade em uma instituição de grande reconhecimento.

Aos meus pais, por todos os esforços e abnegações ao longo da vida, para que eu e minha irmã tivéssemos uma educação de qualidade. A confiança deles de que este era o caminho para uma vida melhor foi essencial para chegar aonde estou hoje, e sou muito grata por isso.

Agradeço também à minha família, que junto com meus pais, sempre me deu o suporte necessário.

Ao meu marido, por me ajudar a conciliar o mestrado com outras responsabilidades e por me apoiar nos momentos mais difíceis, mostrando-me que eu era capaz de seguir em frente.

Aos amigos, sou eternamente grata por estarem sempre ao meu lado, oferecendo apoio, orientação e confiança nos momentos em que mais precisei.

Finalmente, agradeço a empresa em que trabalho, por me oportunizar o desenvolvimento de minha experiência profissional, que foi essencial para a pesquisa, e por viabilizar este trabalho ao fornecer os dados necessários para a pesquisa. Agradeço também às minhas lideranças pelo suporte e compreensão durante todo o período do mestrado.

RESUMO

O comércio eletrônico necessita da organização e representação precisa dos produtos comercializados em plataformas digitais, a fim de oportunizar a busca, recuperação e seleção dos produtos pelos consumidores. Neste contexto, é possível observar a possibilidade de utilizar os processos e instrumentos da área de Organização e Representação do Conhecimento e da Informação para representar produtos do varejo. Portanto, esta pesquisa visa responder a seguinte pergunta: como definir um modelo de representação de produtos do varejo, a partir da perspectiva da área de Organização e Representação do Conhecimento e da Informação, no contexto do comércio eletrônico? O referencial teórico abrange conceitos fundamentais da área de Organização e Representação do Conhecimento e da Informação, como Sistemas de Organização do Conhecimento (SOCs); teorias que fundamentam a elaboração de SOCs, como Teoria da Classificação Facetada, Teoria do Conceito e teorias da Terminologia; além de Catalogação e Metadados. O estudo apresenta abordagem qualitativa, utilizando a Análise de Domínio como metodologia para elaboração de SOCs. O corpus de pesquisa é composto por descrições de produtos da categoria de biscoitos de um marketplace B2B real. O resultado é a proposição de um modelo para representação de produtos do varejo, em apresentação sistemática e formal, detalhando o percurso metodológico para sua elaboração. Este trabalho contribui para a área à medida que amplia a compreensão da Ciência da Informação, especialmente no âmbito da área de Organização e Representação do Conhecimento e da Informação, ao propor um percurso metodológico para a elaboração de SOCs em um contexto não convencional. Além disso, o estudo favorece a prática da área, fornecendo ferramentas tangíveis e aplicáveis com potencial de melhorar a organização e representação de informações sobre produtos no contexto do comércio eletrônico.

Palavras-chave: Organização e Representação da Informação e do Conhecimento; Sistemas de Organização do Conhecimento (SOC); Análise de Domínio; comércio eletrônico; varejo.

ABSTRACT

E-commerce requires the organization and accurate representation of products sold on digital platforms, in order to enable consumers to search for, retrieve and select products. In this context, it is possible to observe the possibility of using processes and tools from the field of Information and Knowledge Organization and Representation to represent retail products. Therefore, this research aims to answer the following question: how can a representation model of retail products be defined from the perspective of the area of Information and Knowledge Organization and Representation, in the context of e-commerce? The theoretical framework covers fundamental concepts in the field of Information and Knowledge Organization and Representation, such as Knowledge Organization Systems (KOS); theories that underpin the development of KOS, such as Faceted Classification Theory, Concept Theory and Terminology theories; as well as Cataloging and Metadata. The study presents a qualitative approach, using Domain Analysis as a methodology for developing KOS. The research corpus consists of product descriptions from the biscuits category from a real B2B marketplace. The result is the proposition of a model for representing retail products, in a systematic and formal presentation, detailing the methodological path for its elaboration. This work contributes to the area by broadening the understanding of Information Science, especially in the area of Information and Knowledge Organization and Representation, by proposing a methodological path for the elaboration of KOS in an unconventional context. Furthermore, the study favors the practice of the area, providing tangible and applicable tools with the potential to improve the organization and representation of product information in the context of e-commerce.

Keywords: Information and Knowledge Organization and Representation; Knowledge Organization System (KOS); Domain Analysis; e-commerce; retail.

LISTA DE FIGURAS

Figura 1 –	Triângulo Conceitual de Dahlberg.....	52
Figura 2 –	Processo de conceitualização.....	54
Figura 3 –	Dimensões do mix de produtos.....	112
Figura 4 –	Processo de gerenciamento por categorias.....	115
Figura 5 –	Árvore de decisão para desodorantes.....	117
Figura 6 –	Exemplo de estrutura mercadológica.....	117
Figura 7 –	Árvore de decisão para biscoitos.....	129
Figura 8 –	Estrutura GPC.....	134
Figura 9 –	Atributos GPC.....	134
Figura 10 –	GPC (Classe Biscoitos).....	136
Figura 11 –	Estrutura do GoodRelations.....	138
Figura 12 –	Representação de produtos e assuntos na eClassOWL.....	140
Figura 13 –	Estrutura taxonômica da categoria de biscoitos.....	142
Figura 14 –	Classe <i>food product</i> e suas propriedades.....	143
Figura 15 –	Classe <i>Cookie Food Product</i> na FoodOn.....	144
Figura 16 –	Comparação taxonomia FPC e Google.....	145
Figura 17 –	Exemplo de aplicação de categoria do Google.....	146
Figura 18 –	Processo de categorização.....	155
Figura 19 –	Esquema conceitual.....	156
Figura 20 –	Comparação de taxonomias de referência.....	157
Figura 21 –	Aplicação de metacategorias semânticas.....	160
Figura 22 –	Estrutura taxonômica da faceta <i>Treatment Applied</i>	172
Figura 23 –	Esquema dos procedimentos metodológicos aplicados na pesquisa.....	173
Figura 24 –	Taxonomia de tipos de biscoitos.....	175
Figura 25 –	Taxonomia final de tipologia de biscoitos.....	179
Figura 26 –	Apresentação sistemática do modelo de representação.....	187
Figura 27 –	Exemplo de aplicação do modelo: apresentação sistemática.....	202

LISTA DE QUADROS

Quadro 1 –	Quantificação de produtos por segmento.....	87
Quadro 2 –	Procedimentos e Objetivos específicos.....	89
Quadro 3 –	Abordagens de AD adotadas na pesquisa.....	100
Quadro 4 –	Extração de termos de descrição.....	148
Quadro 5 –	Definições dos termos extraídos.....	149
Quadro 6 –	Processo de conceitualização.....	152
Quadro 7 –	Caracterização das categorias.....	154
Quadro 8 –	Variações terminológicas das classes.....	163
Quadro 9 –	Controle de vocabulário das características.....	164
Quadro 10 –	Variações terminológicas das características.....	165
Quadro 11 –	Padrão de metadados.....	168
Quadro 12 –	Aplicação do padrão de metadados.....	172
Quadro 13 –	Comparativo entre hierarquia de necessidades e níveis de estrutura mercadológica.....	178
Quadro 14 –	Definições das classes da taxonomia.....	180
Quadro 15 –	Definições das metacategorias.....	188
Quadro 16 –	Definições das características.....	190
Quadro 17 –	Variações terminológicas das características.....	194
Quadro 18 –	Padrão de metadados adotado no modelo.....	196
Quadro 19 –	Exemplo de aplicação do modelo: extração dos termos e conceitualização.....	201
Quadro 20 –	Exemplo de aplicação do modelo: ficha de metadados.....	203
Quadro 21 –	Objetivos específicos e Resultados.....	204

SUMÁRIO

1 INTRODUÇÃO	12
1.1 JUSTIFICATIVA	16
1.2 OBJETIVOS	19
1.2.1 Objetivo geral	19
1.2.2 Objetivos específicos	19
2 FUNDAMENTAÇÃO TEÓRICA	20
2.1 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO	21
2.1.1 Sistemas de Organização do Conhecimento	24
2.1.1.1 Listas e Anéis de sinônimos	26
2.1.1.2 Taxonomia.....	27
2.1.1.3 Tesouro	33
2.1.1.4 Ontologia	39
2.1.2 Bases teóricas para desenvolvimento de Sistemas de Organização do Conhecimento	44
2.1.2.1 Teoria da Classificação Facetada	45
2.1.2.2 Teoria do Conceito	52
2.1.2.3 Terminologia.....	62
2.2 CATALOGAÇÃO E METADADOS.....	68
2.2.1 Catalogação	69
2.2.2 Metadados	74
3 METODOLOGIA E PROCEDIMENTOS	81
3.1 CONTEXTUALIZAÇÃO DA PESQUISA	81
3.2 CARACTERIZAÇÃO DA PESQUISA.....	84
3.3 PROCEDIMENTOS METODOLÓGICOS	85
3.3.1 Planejamento e coleta de dados	85
3.3.2 Preparação dos dados	87
3.3.3 Tratamento dos dados	88
3.3.4 Análise dos dados	88
3.3.5 Apresentação dos resultados	90
3.4 ANÁLISE DE DOMÍNIO	90
3.4.1 Definição de Domínio	94
3.4.2 Metodologia em Análise de Domínio	97

4 RESULTADOS	102
4.1 DOMÍNIO DE PRODUTOS DO VAREJO	102
4.1.1 Varejo	103
4.1.2 Produto	107
4.1.3 Biscoitos	125
4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento	130
4.1.4.1 Schema.org	130
4.1.4.2 GS1	133
4.1.4.3 GoodRelations.....	137
4.1.4.4 Product Type Ontology	139
4.1.4.5 eClassOWL	140
4.1.4.6 LanguaL	141
4.1.4.7 FoodOn	143
4.1.4.8 Google e Facebook	145
4.2 DESENVOLVIMENTO DO MODELO DE REPRESENTAÇÃO	147
4.3 MODELO DE REPRESENTAÇÃO DE PRODUTOS DO VAREJO	175
5 CONSIDERAÇÕES FINAIS	206
REFERÊNCIAS	210
APÊNDICE A – Descrições dos produtos	226
APÊNDICE B – Termos extraídos das descrições dos produtos	226
APÊNDICE C – Definições dos termos extraídos	226
APÊNDICE D – Compilado de metadados de referência	226
APÊNDICE E – Modelo de representação completo em Excel	226

1 INTRODUÇÃO

O varejo compreende todas as atividades relacionadas à venda de bens ou serviços diretamente ao consumidor final, para uso pessoal (Kotler; Keller, 2012). O setor varejista é um dos mais relevantes na economia brasileira, representando 22,9% do PIB brasileiro (Sociedade Brasileira de Varejo e Consumo, 2022).

A entrega de um produto ao cliente ocorre por meio da cadeia logística, que vai desde a produção do item, passando pelo seu armazenamento, expedição, transporte, recebimento, exposição e comercialização (Rocha; Mola, 2018). Os estabelecimentos varejistas representam o ponto de venda, espaço no qual o consumidor tem contato direto com o produto, tanto bens duráveis quanto bens de consumo, sendo o último elo da cadeia logística (Rocha; Mola, 2018; Pereira; Barros, 2017).

O produto é tudo que pode ser oferecido a um mercado para satisfazer alguma necessidade ou desejo, podendo incluir bens físicos, serviços, além de experiências, eventos, lugares, entre outros (Kotler; Keller, 2012). Nesta pesquisa, o produto é considerado como um bem material, e não como um serviço.

O varejista pode estar estabelecido tanto em uma loja física quanto em lojas virtuais, na *internet*. A década de 1990 marcou o início do *e-commerce* (comércio eletrônico), quando surgiram as primeiras lojas virtuais em que empresas ofereciam seus produtos para compra pelos consumidores, por meio de uma plataforma digital na *web*. Neste tipo de ambiente, a maneira como o consumidor compra é diferente do que em uma loja física. Não há corredores pelos quais o consumidor percorre e pode ser atraído por alguma marca ou embalagem. O consumidor não tem contato direto com o produto que está comprando. O contato do consumidor com o produto é limitado à dimensão cognitiva, ou seja, a interação com o produto é delimitada pelas informações que são fornecidas sobre ele (Böse, 2016).

O conjunto das informações que descrevem o produto vendido em uma plataforma digital compõe a representação deste produto. Aqui encontra-se o ponto de convergência entre os produtos do varejo comercializados por meio do comércio eletrônico e a área de Organização e Representação do Conhecimento e da Informação. Esta área contempla as teorias e práticas para construção de sistemas conceituais que organizam e representam o conhecimento de um domínio específico (Organização e Representação do Conhecimento). Estes sistemas são utilizados na organização e representação de objetos informacionais, por meio de processos de

catalogação, classificação e indexação (Organização e Representação da Informação).

De modo geral, todo o contato que temos com o mundo virtual se dá por meio de representação, pois não estabelecemos um contato direto com uma entidade em si, mas com a sua representação exposta em tela. Na área da Ciência da Informação (CI), pode-se observar este cenário nas bibliotecas e documentos digitais, por exemplo. Neste contexto, apesar de haver uma mudança no suporte que carrega a informação, ainda se trata de um conteúdo informacional, como um texto, imagem, som etc., trazendo em si elementos temáticos e descritivos. Portanto, mesmo no ambiente digital, é possível observar a possibilidade de utilizar os processos e instrumentos da área de Organização e Representação do Conhecimento e da Informação de forma tradicional, a partir da perspectiva dicotômica de representação temática (assuntos) e representação descritiva (elementos descritivos).

Já no caso dos produtos do varejo, apesar de também carregarem informação sobre si mesmos, seu principal valor está no mercado, suprimindo uma demanda de consumo. Ao contrário de um documento ou um livro, em que “consumimos” a informação contida neles; no caso dos produtos, o que é consumido é o próprio produto, para finalidades específicas. Por exemplo, um biscoito para alimentação, sabão em pó para limpeza de roupas, ou uma escova de dentes para a higiene oral.

Apesar destas diferenças, é possível vislumbrar as contribuições que a área de Organização e Representação do Conhecimento e da Informação pode trazer para a organização e representação de produtos do varejo em plataformas digitais.

A recuperação da informação está diretamente relacionada a organização e representação da informação. No contexto do *e-commerce*, o consumidor acessa páginas de pesquisa ou de lojas na *web*, com objetivo de localizar determinado produto que está procurando ou algum produto com as determinadas características que ele está buscando, para atender às suas necessidades de consumo. Neste cenário, o consumidor utiliza o menu da página, para navegar entre as categorias de produtos, ou pesquisa no mecanismo de busca diretamente por termos relacionados ao tipo de produto ou seus atributos. O resultado da recuperação da informação depende da forma pela qual as informações sobre os produtos foram organizadas e representadas na plataforma digital.

Por exemplo, no caso da navegação por categorias, se o usuário estiver buscando um liquidificador e houver categorias como “cozinha” e “eletroportáteis”,

gerará confusão e dificuldade para o usuário localizar o produto que está procurando. Neste sentido, os fundamentos da Organização do Conhecimento podem ser aplicados para melhorar a forma como os produtos são organizados. Por meio da Teoria do Conceito é possível identificar as principais características que formam o conceito de determinado produto, e partir disso, identificar as similaridades e diferenças deste produto em relação aos demais. Utilizando a Teoria da Classificação Facetada, é possível, a partir da identificação das características do conceito, definir qual característica é mais relevante para o usuário no contexto em questão, que poderá ser utilizada como critério de subdivisão para a formação de estruturas hierárquicas, evitando a ambiguidade gerada pelo exemplo citado.

Outra situação que pode ser encontrada nestes contextos, é a falta de controle de vocabulário na representação dos produtos. Por exemplo, na categoria de “luminárias”, pode haver um filtro para “tensão elétrica” trazendo opções como “bivolt” e “110/220V”, que possuem o mesmo significado, fazendo com que cada uma destas opções não recupere todos os produtos que apresentam esta característica. Neste sentido, a utilização de Sistemas de Organização do Conhecimento (SOCs) que possibilitem o controle de vocabulário são importantes para a consistência da representação da informação, e a conseqüente melhoria na sua recuperação.

Essas situações destacam a importância da organização e representação da informação para garantir que os consumidores encontrem e acessem os produtos buscados. Além disso, observa-se o impacto comercial causado pela forma como a informação é organizada e representada: quando o consumidor não encontra o que procura ou a descrição do produto não fornece todas as informações necessárias para a tomada de decisão, ele pode abandonar a compra, impactando diretamente o negócio.

Esta pesquisa busca, portanto, trazer o conhecimento e as práticas da CI, mais especificamente da área de Organização e Representação do Conhecimento e da Informação, para serem empregados no contexto do *e-commerce*, mas ainda mantendo o mesmo objetivo principal: fazer com que o usuário encontre os melhores resultados para a sua necessidade de busca. E isto, no contexto comercial, favorece o crescimento e melhoria de empresas e negócios, evidenciando o impacto social e econômico desta pesquisa.

A partir destes elementos, constitui-se o problema de pesquisa: como definir um modelo de representação de produtos do varejo, a partir da perspectiva da área

de Organização e Representação do Conhecimento e da Informação, no contexto do comércio eletrônico?

Para isso, este trabalho buscou suporte nos fundamentos teóricos e aplicados desta área, estudando os principais tipos de SOCs, entendendo suas características e de que forma poderiam ser utilizados neste contexto não tradicional dentro da CI.

Além disso, compreendendo que mais importante do que os instrumentos utilizados, são os seus fundamentos, para pensar em novas formas de aplicação, buscou-se fundamentação na Teoria da Classificação Facetada, que já vem se mostrando bastante adequada à organização da informação em ambientes digitais; na Teoria do Conceito, para compreender as características dos produtos; bem como nas teorias da Terminologia, visto que os produtos do varejo estão inseridos em um domínio do conhecimento que apresenta uma linguagem especializada. Devido aos elementos descritivos serem mais relevantes na representação de produtos, buscou-se também fundamentação nos princípios da catalogação e em metadados, que são utilizados para representar formalmente objetos digitais em sistemas de informação.

A pesquisa é realizada em um contexto real, sendo este a instituição em que a pesquisadora atua profissionalmente, devido à facilidade de acesso aos dados e familiaridade com o domínio de conhecimento. São dados referentes a produtos do varejo alimentar comercializados entre empresas, de indústrias para minimercados localizados em São Paulo e região metropolitana, por meio de plataforma de comércio eletrônico.

O recorte de pesquisa é composto por uma categoria da taxonomia já existente na empresa. Foi selecionada a categoria de biscoitos, pois, para o contexto em questão, os produtos de mercearia seca, como biscoitos, são priorizados na empresa, devido a maior praticidade logística em comparação com produtos perecíveis, por exemplo. Além disso, o Brasil movimentou R\$ 29 bilhões em 2022 em valor e 1,5 milhões de toneladas em volume de vendas de biscoitos. Pesquisa realizada pela Indexsa, revelou que o biscoito é um alimento que tem penetração de 99,7% nos lares brasileiros. Estes números demonstram a importância desta categoria de produtos alimentícios no mercado e nos hábitos de consumo dos brasileiros (Biscoito [...], 2023).

Quanto à metodologia, esta pesquisa emprega abordagem qualitativa, buscando compreender os elementos que constituem o objeto que será estudado,

visando seu aprofundamento por meio de interpretação e estruturação dos dados coletados (Michel, 2015).

A Análise de Domínio (AD) foi utilizada como base metodológica com objetivo de compreender o domínio de conhecimento de produtos do varejo, por meio de revisão de literatura na área, configurando o corpus textual especializado da pesquisa, identificando os principais conceitos e como eles se relacionam no domínio de conhecimento. Além disso, buscou-se na literatura da área, modelos de SOCs e padrões de metadados utilizados para a organização e representação de informações sobre produtos. O corpus textual especializado foi utilizado para análise e validação dos dados coletados no corpus de pesquisa.

O corpus de pesquisa é formado por amostra do recorte estabelecido, sendo este a categoria de biscoitos. Foi realizada análise qualitativa com os dados coletados de base de dados com descrições dos produtos cadastrados na plataforma de comércio eletrônico do contexto em questão. Os procedimentos para análise foram organizados nas seguintes etapas: planejamento e coleta de dados, preparação dos dados, tratamento dos dados, análise dos dados e apresentação dos resultados, em que será apresentada proposta de modelo de representação de produtos do varejo, em uma estrutura organizada de maneira sistemática e formal.

O percurso metodológico delineado no capítulo de metodologia teve sua execução explicada detalhadamente no capítulo de resultados, de forma que promova a reutilização desse percurso metodológico com objetivo de representar outros tipos de produtos, inclusive em outros contextos diferentes do *e-commerce*, abordado neste trabalho. Desta forma, esta pesquisa busca apresentar como resultado, além do modelo de representação para produtos do varejo, um guia detalhado que pode ser seguido para o desenvolvimento de outros modelos de representação de produtos.

1.1 JUSTIFICATIVA

No estudo do domínio do conhecimento do varejo, especialmente considerando os processos e ferramentas adotadas por esta área para o gerenciamento de categorias e informações sobre produtos, é possível observar que existe no mercado varejista a compreensão da importância da informação e da organização de produtos para a melhoria do desempenho de vendas e da gestão no varejo. Há processos, ferramentas e estratégias na área da gestão do varejo e de produtos para lidar com

este cenário, e que perpassam por conceitos da área de Organização e Representação do Conhecimento e da Informação, como taxonomia, atributos de produtos, catálogos, entre outros.

Portanto, revela-se a oportunidade para contribuição de um estudo interdisciplinar entre estas áreas, visando a melhoria do processo de organização da informação de produtos do varejo, bem como a possibilidade de ampliação da atuação profissional e da pesquisa para a área de Organização e Representação do Conhecimento e da Informação.

Em levantamento realizado na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), entre 1996 (ano de criação da primeira loja virtual no Brasil) e 2023, foram localizados vinte e seis (26) artigos, buscando pelo termo “comércio eletrônico”. Destes vinte e seis artigos, apenas dois apresentavam relação entre comércio eletrônico e algum aspecto da área de Organização e Representação do Conhecimento e da Informação. Os dois artigos foram citados neste trabalho: “Taxonomias navegacionais em sítios de comércio eletrônico: critérios para avaliação”, de Raphael da Silva Cavalcante e Marisa Bräscher; e “Taxonomia para comércio eletrônico: diferentes perspectivas em *front e back end*”, de Isadora dos Santos Garrido Steimer e Charley dos Santos Luz. Pesquisando na BRAPCI, no mesmo período, pelos termos “produto” e “varejo”, utilizando o operador booleano AND na busca, não houve retorno de nenhum resultado. Portanto, verifica-se a lacuna de estudos relacionados a produtos do varejo em comércio eletrônico e a área de Organização e Representação do Conhecimento e da Informação.

Do ponto de vista da atuação profissional, Alvarenga (2001) acredita que os bibliotecários deveriam constituir uma classe profissional com atuação mais marcante, na tarefa de organizar os objetos digitais no ambiente da *internet*. Porém, na realidade, os profissionais da área da CI ainda se mantêm tímidos, apresentando contribuições aquém de suas possibilidades. Steimer e Luz (2015) declaram que, para uma vaga em comércio eletrônico de varejo, o profissional da informação deve estar preparado para lidar com diferentes equipes, como engenharia de *software*, pesquisa, produto e negócio, além de lidar com legados informacionais de *softwares* já utilizados e em desenvolvimento.

A concepção desta pesquisa parte do interesse em analisar como a área da Organização e Representação do Conhecimento e da Informação pode contribuir, por

meio de sua teoria e prática com a melhoria da organização e representação de informações sobre produtos do varejo em plataformas digitais de comércio eletrônico. Do ponto de vista particular, este trabalho se justifica pelo interesse profissional da pesquisadora que atua com cadastro de produtos para um comércio eletrônico B2B (*Business-to-Business*). Em sua experiência, foi possível observar dois pontos diretamente ligados à área de Organização e Representação do Conhecimento e da Informação: a taxonomia (conhecida também como estrutura mercadológica) e os metadados descritivos dos produtos.

O desenvolvimento de um modelo de representação de produtos do varejo com base nas teorias e práticas da área da Organização e Representação do Conhecimento e da Informação, pretende trazer para a área da CI, elementos metodológicos que contribuam para a área instrumentalizar seus processos em novos contextos digitais de aplicação, pelos profissionais da informação. Do mesmo modo, a construção de um modelo de representação pretende disponibilizar ferramentas próprias da CI para lidar com problemas particulares da organização e representação de informação sobre produtos do varejo, como a recuperação da informação, o que pode impactar diretamente nas vendas e no desempenho dos negócios.

Este trabalho está organizado nos seguintes capítulos: Fundamentação Teórica, Metodologia e Procedimentos, Resultados e Considerações Finais. O capítulo de Fundamentação teórica apresenta os fundamentos da área de Organização e Representação do Conhecimento e da Informação, incluindo seus principais sistemas, processos e bases teóricas para desenvolvimento de seus instrumentos, e da área de Catalogação, incluindo a utilização de metadados como meio para representar formalmente objetos informacionais. O capítulo de Metodologia e Procedimentos demonstra como serão compostos o corpus textual especializado e o corpus de pesquisa, e de que forma será realizada a análise dos dados e apresentação dos resultados. O capítulo de Resultados apresenta o corpus textual, especializado formado por meio da Análise de Domínio, e descreve o processo de análise dos dados do corpus de pesquisa até chegar no resultado final, o modelo de representação. Por fim, o capítulo de Considerações Finais reflete sobre os principais pontos analisados e discutidos na pesquisa, aponta as principais contribuições promovidas por este trabalho, além de indicar possibilidades para estudos futuros.

1.2 OBJETIVOS

Nas seções subsequentes são apresentados o objetivo geral e objetivos específicos desta pesquisa.

1.2.1 Objetivo geral

Propor modelo de representação de produtos do varejo, a partir da perspectiva da área de Organização e Representação do Conhecimento e da Informação, no contexto do comércio eletrônico.

1.2.2 Objetivos específicos

- a) Identificar e analisar os atributos relevantes dos produtos que compõem o recorte da pesquisa;
- b) Identificar e relacionar categorias de representação a partir destes atributos;
- c) Estabelecer controle terminológico dos termos adotados nas categorias;
- d) Estruturar modelo de representação para o recorte estudado.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta pesquisa, o objeto informacional que será descrito é o produto comercializado no varejo. Portanto, para o desenvolvimento de modelo de representação do produto, serão utilizados como aportes teóricos, a Organização e Representação do Conhecimento, com o objetivo de conceituar o que é o produto, quais são os seus tipos, características e seus relacionamentos, e a Catalogação e Metadados, com objetivo de identificar quais os elementos descritivos sobre o produto devem ser considerados em um modelo, para que o produto seja representado e, conseqüentemente, recuperado.

Nesta seção serão apresentados os fundamentos teóricos a partir dos quais serão analisados os dados coletados, buscando atingir os objetivos desta pesquisa. O referencial teórico da dissertação está dividido em duas partes: Organização e Representação do Conhecimento, e Catalogação e Metadados. A primeira parte apresenta a área de Organização e Representação do Conhecimento, com foco nos SOCs e nas bases teóricas para seu desenvolvimento. A segunda parte apresenta fundamentos da catalogação e padrões de metadados, objetivando a organização e representação de recursos informacionais digitais.

Desta forma, a fundamentação teórica servirá para identificar os principais aspectos de instrumentos de organização e representação do conhecimento e da informação para que, com base na realidade do domínio do conhecimento em questão, seja possível elencar os elementos necessários para o desenvolvimento de um modelo para a organização e representação de informações sobre produtos do varejo, considerando as necessidades do contexto específico.

Antes de iniciar a fundamentação teórica, é importante destacar a diferença entre Organização e Representação do Conhecimento e a Organização e Representação da Informação, devido à própria diferença entre o que é conhecimento e o que é informação.

Conhecimento é a certeza de alguém sobre um fato, que é elaborado pela própria reflexão pessoal (Dahlberg, 2006). “[...] está relacionado com os aspectos cognitivos que ocorrem na mente humana e envolvem os processos mentais de captação, assimilação, associação e também de construção, desconstrução e reconstrução de conceitos.” (Lima; Alvares, 2012, p. 24).

Já a informação é este conhecimento registrado de alguma forma, por meio da linguagem, em texto ou imagens, em papel ou digitalmente. Através do conhecimento registrado, o indivíduo pode novamente alterar e criar novos conhecimentos, que voltam a ser registrados como informação. Este processo dialoga com a cultura do indivíduo, seus valores e princípios. Portanto, “[...] 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.” (Lima; Alvares, 2012, p. 25). A CI trata da organização e representação do conhecimento e da informação a partir de sua perspectiva social, focando na informação e conhecimento que são compartilhados.

Para Dahlberg (2006) a organização e representação do conhecimento compreende a elaboração de sistemas conceituais, como sistemas de classificação, tesouros etc., que organizam o conhecimento de um domínio específico através da organização de seus conceitos, gerando uma representação conceitual deste domínio.

Já a organização e representação da informação compreende a utilização destes sistemas conceituais para a descrição de objetos informacionais (DAHLBERG, 2006). Para Bräscher e Café (2010), a organização e representação da informação envolve a descrição física e de conteúdo de objetos informacionais. Desta forma, a organização e representação da informação abrange a descrição temática do objeto informacional, utilizando os instrumentos criados a partir da organização e representação do conhecimento, por meio dos processos de indexação e classificação, mas também envolve a descrição dos aspectos físicos do objeto informacional, relacionados com os processos de catalogação.

2.1 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

A Organização do Conhecimento (OC) é um campo de estudo, ensino e prática, principalmente afiliado a Biblioteconomia e a CI. De acordo com Hjørland (2016b, p. 475, tradução nossa): “Organização do Conhecimento é sobre descrever, representar,

arquivar e organizar documentos, representações de documentos, assuntos e conceitos tanto por humanos quanto por programas de computador.”¹

O termo “Organização do Conhecimento”, em inglês, “*Knowledge Organization*”, foi introduzido pelo periódico “*Knowledge Organization (KO)*”, intitulado à época como “*International Classification*”, criado em 1974 por Ingetraut Dahlberg (DAHLBERG, 2006). Segundo Dahlberg (2006, p. 12, tradução nossa), a OC “[...] compreende os objetos e atividades da teoria do conceito, classificação e indexação e representação do conhecimento”².

Em sentido mais restrito, a OC trata sobre a elaboração de SOCs e registros bibliográficos. Já em sentido mais amplo do termo, a OC trata sobre como o conhecimento é organizado em diferentes domínios e como isto pode ser utilizado para a recuperação da informação (Hjørland, 2016b).

De acordo com Hjørland (2016b), os dois principais aspectos da OC são os Processos de Organização do Conhecimento e os SOCs:

Processos de Organização do Conhecimento são, por exemplo, os processos de catalogação, análise de assunto, indexação e classificação por humanos ou computadores. Os Sistemas de Organização do Conhecimento são a seleção de conceitos com indicação de relações semânticas selecionadas, por exemplo, sistemas de classificação, listas de cabeçalho de assunto, tesouros, ontologias e outros sistemas de metadados. (HJØRLAND, 2016b, p. 475, tradução nossa).³

Hjørland (2016b) também apresenta abordagens da OC:

a) abordagens dentro da OC:

- abordagem prática e intuitiva: prioriza o aspecto prático, bem como utiliza o mesmo sistema de classificação para diversas unidades de informação, facilitando a centralização da classificação e indexação. Não é utilizado nenhum tipo de critério baseado em pesquisa, priorizando a intuição;

¹ No original: “KO is about describing, representing, filing and organizing documents, document representations, subjects and concepts both by humans and by computer programs.” (Hjørland, 2016b, p. 475).

² No original: “[...] comprising the objects and activities of [...] concept theory, classification and indexing and knowledge representation” (Dahlberg, 2006, p. 12).

³ No original: “Knowledge organization processes (KOP) are, for example, the processes of cataloging, subject analysis, indexing and classification by humans or computers. Knowledge organization systems (KOS) are selection of concepts with indication of selected semantic relations, for example, classification systems, lists of subject headings, thesauri, ontologies and other systems of metadata.” (Hjørland, 2016b, p. 475).

- abordagem baseada em consenso: trazida por Henry E. Bliss, defende que a classificação deve se basear em consenso científico e educacional. Não contempla interesses e perspectivas conflitantes no mesmo domínio do conhecimento;
- abordagem analítico-facetada: atribuída a S. R. Ranganathan, fundamentada nos princípios lógicos de divisão. Provavelmente a abordagem mais distinta no campo da Biblioteconomia e Ciência da Informação;
- abordagem cognitiva e baseada no usuário: estruturação de SOCs baseada em estudos de usuário. Não é uma hipótese com suporte na área;
- abordagem de análise de domínio/epistemológica: considera que um objeto pode ser classificado a partir de múltiplas perspectivas. Nesta abordagem, o termo Organização do Conhecimento reflete a conexão com as metaciências;

b) abordagens fora da OC:

- abordagem bibliométrica: a Bibliometria é um campo interdisciplinar, que desenvolve técnicas para a produção de mapas bibliométricos, que servem para a recuperação da informação e é uma abordagem complementar da Organização do Conhecimento;
- abordagem de Recuperação da Informação: estudo de relações estatísticas entre termos, documentos e coleções de documentos;
- outras abordagens: a abordagem ontológica fornece especificações formais e definições harmonizadas de conceitos usados para representar o conhecimento de domínios específicos.

Neste trabalho, serão adotadas a abordagem analítico-facetada, visto que a Teoria da Classificação Facetada será utilizada como uma das bases teóricas para o desenvolvimento de SOCs, e a abordagem de análise de domínio, na medida em que a AD servirá como base metodológica para a identificação e análise dos elementos que formam o domínio de produtos do varejo.

A AD abrange os diferentes paradigmas presentes em um domínio, sendo fundamental para a seleção de fontes, definição de conceitos e determinação dos critérios de organização da estrutura conceitual. Dessa forma, a AD, quando adotada como metodologia para o desenvolvimento de SOCs, complementa a classificação

facetada, ao fornecer os elementos empíricos necessários para a classificação de um domínio – elementos esses que a metodologia da análise facetada, por si só, não oferece (Hjørland; Barros, 2024).

O objeto da área da OC é o conhecimento. De acordo com Dahlberg (2006), “conhecimento” neste contexto é entendido no sentido de “conhecido” e “organização” é compreendida no sentido da atividade de construir algo de acordo com um plano. Segundo Dahlberg (2006, p. 12, tradução nossa), o conhecimento é: “[...] a certeza subjetiva e objetiva razoavelmente bem fundamentada de alguém sobre a existência de um fato ou assunto. Este conhecimento não é transferível, só podendo ser elaborado pela própria reflexão pessoal de alguém.”⁴

O processo de criação de conhecimento na mente humana passa pela classificação, pois ao nos depararmos com um novo conceito a ser aprendido, o comparamos com conceitos já conhecidos em nossa mente, buscando agrupar conceitos com características semelhantes, de acordo com algum critério. Dessa forma, “[...] a classificação transforma impressões isoladas em conceitos reconhecíveis.” (Schiessl; Shintaku, 2012, p. 59). Este conhecimento criado pode ser transferido no espaço e no tempo por meio da linguagem e pode ser comunicado por palavras ou sinais falados e escritos (Dahlberg, 2006).

Na vida diária, utilizamos a linguagem natural para nos comunicarmos e transmitirmos conhecimento. Porém, para a comunicação em áreas especializadas do conhecimento, são criadas as chamadas linguagens especiais, conhecidas também por linguagens documentárias (Dahlberg, 1978b). Elas são formadas por uma estrutura conceitual própria que representa aquele domínio do conhecimento. Atualmente, são denominadas como SOCs, pois agregam elementos incorporados nas inovações tecnológicas da era digital (Carlan; Medeiros, 2011).

2.1.1 Sistemas de Organização do Conhecimento

Os SOCs são sistemas conceituais estruturados que contemplam termos, definições, relacionamentos e propriedades dos conceitos (Carlan; Medeiros, 2011). De acordo com Schiessl e Shintaku (2012, p. 49), “[...] são representações do

⁴ No original: “Knowledge is the subjectively and objectively fairly well-founded certainty of somebody about the existence of a fact or a matter. This knowledge is not transferable, it can only be elaborated by somebody's own personal reflection.” (Dahlberg, 2006, p. 12).

conhecimento, que por sua vez são modelos de abstração do mundo real, construídos para determinada finalidade”.

No contexto da CI, trata-se de instrumentos criados para a organização e representação de um domínio do conhecimento, propiciando a organização e representação de recursos informacionais, em qualquer suporte, na perspectiva da Organização e Representação da Informação, através dos processos de catalogação, indexação e classificação, em um Sistema de Recuperação da Informação (SRI) (Lima, 2020).

O termo Sistema de Organização do Conhecimento, em inglês *Knowledge Organization System* (KOS), foi inicialmente utilizado por Hodge (2000), abrangendo todos os tipos de esquemas de organização da informação e gerenciamento do conhecimento. O termo também foi empregado por Soergel (2001), quando analisou as normas dos instrumentos de Organização do Conhecimento. De acordo com Carlan e Medeiros (2011, p. 55): “[...] pode-se dizer que sistema de organização do conhecimento é uma denominação nova para as linguagens documentárias que agregam elementos incorporados nas inovações tecnológicas da era digital.”

Quanto à estrutura, os SOCs variam de um esquema simples até a um esquema mais complexo (multidimensional). Dentre suas principais funções estão a eliminação da ambiguidade, controle de sinônimos ou equivalentes e estabelecimento de relacionamentos semânticos entre conceitos (Carlan; Medeiros, 2011).

Em relação ao desenvolvimento dos SOCs, é necessária uma: “[...] análise das necessidades dos usuários dos sistemas; a identificação do tipo de SOC apropriado e o desenvolvimento do hardware e do software adequado à arquitetura de rede, sua integração e manutenção.” (Carlan; Medeiros, 2011, p. 55).

Os SOCs abrangem sistemas de classificação, listas de autoridade, anéis de sinônimos, listas de cabeçalhos de assunto, taxonomias, tesouros, ontologias entre outros. A norma ANSI/NISO Z39.19:2005 (National Information Standards Organization, 2010) considera quatro tipos diferentes de SOC, sendo eles as listas, anel de sinônimos, taxonomias e tesouros. Estes instrumentos se assemelham e se diferenciam pelo nível de estruturação semântica de seus elementos (Lara, 2015).

Zeng (2008) classifica os diferentes tipos de SOC em quatro classes, de acordo com o nível de complexidade de suas estruturas e funções principais, sendo estas: listas de termos (listas de seleção, dicionários, glossários e anéis de sinônimos) que tratam do controle de ambiguidade e de sinônimos; modelos semelhantes a

metadados (controle de autoridade, diretórios e *gazetteers*) que adicionalmente apresentam relações hierárquicas; classificação e categorização (cabeçalhos de assunto, esquemas de categorização, taxonomias e esquemas de classificação); e modelos de relacionamentos (tesauros, redes semânticas e ontologias) que adicionalmente apresentam relações associativas e propriedades.

Já Souza, Tudhope e Almeida (2012) apresentam uma divisão semelhante, mas trazendo mais tipos de SOCs em cada categoria. Eles classificam os SOCs em textos não estruturados, como resumos; listas de termos/conceitos; estruturas de conceitos, relacionamentos e *layout*; e estruturas de conceitos e relacionamentos, abrangendo taxonomias, tesauros e ontologias.

Nas próximas seções, segue uma breve explanação sobre os principais tipos de SOCs considerados relevantes para o desenvolvimento desta pesquisa.

2.1.1.1 Listas e Anéis de sinônimos

Na classificação dos SOCs, dos mais simples aos mais complexos, as listas e os anéis de sinônimos encontram-se na classe que agrupa SOCs considerados como listas de termos.

As listas são um conjunto de termos organizados em ordem alfabética, ou em alguma outra ordem relevante no contexto de aplicação, e servem para eliminar ambiguidades em um vocabulário. De acordo com Zeng (2008), as principais características das listas são:

- a) todos os termos elencados na lista devem fazer parte do mesmo grupo ou classe de itens;
- b) não pode haver sobreposição de significado entre os termos; e
- c) devem apresentar o mesmo nível de granularidade.

Já os anéis de sinônimos servem para o controle de sinônimos e variações terminológicas especificamente para recuperação da informação, não sendo utilizados na indexação, visto que nestes instrumentos não há um termo autorizado ou preferido. Podem ser usados quando um conceito é descrito por diversos termos sinônimos ou quase-sinônimos, garantindo que um conjunto de informações seja recuperado independentemente por qual destes termos o usuário tenha realizado sua pesquisa (Zeng, 2008).

Estes SOCs configuram-se como alguns dos tipos mais simples de SOCs, apresentando os conceitos em uma única dimensão, sem relacionamentos entre os conceitos. Porém, apesar de sua simplicidade, fornecem soluções para eliminação de ambiguidade e controle de sinônimos que podem ser aplicadas em sistemas de informação de forma mais simples, mas eficaz, dependendo da necessidade.

2.1.1.2 Taxonomia

As taxonomias são estruturas classificatórias, construídas de forma sistemática e organizadas por semelhanças e diferenças entre características do objeto ou fenômeno em um domínio específico, em que são divididos em classes, que são subdivididas em subclasses sucessivamente, do geral ao mais específico, hierarquicamente (Carlan; Medeiros, 2011). É uma forma de representar o conhecimento, disposta metaforicamente como uma árvore invertida, em formato *top-down*. A estrutura se inicia com uma raiz, sendo uma classe mais genérica (superclasse), e se ramifica em classes mais específicas (subclasses), possuindo características extras em relação à respectiva superclasse, devido ao princípio da herança, formando uma estrutura de generalização-especificação, ou seja, gênero-espécie (Schiessl; Shintaku, 2012).

A estrutura hierárquica da taxonomia não é composta apenas da relação gênero-espécie, mas também pode ser formada por relações de todo-parte, em que a superclasse apresenta um aspecto do todo, enquanto a subclasse apresenta aspecto de uma parte do todo (Schiessl; Shintaku, 2012). A relação partitiva acomoda uma conveniência na organização da informação, pois neste tipo de estrutura, a subclasse não herda as características da superclasse, mas indica que é uma parte integrante do todo representado na superclasse (Schiessl; Shintaku, 2012).

Na acepção clássica do termo, o conceito de taxonomia está intimamente relacionado à classificação e hierarquização, sendo legítimo considerar primariamente as taxonomias como ferramentas destinadas à classificação, a partir de uma estrutura hierarquizada (Cavalcante; Bräscher, 2014).

Mais recentemente, a norma ANSI/NISO Z39.19:2005 (National Information Standards Organization, 2010), em sua definição de taxonomia, destacou o aspecto de controle de vocabulário e a possibilidade de polihierarquia em sua estrutura. Quanto ao controle de vocabulário, a norma ANSI/NISO Z39.19:2005 afirma que: “[...]”

o controle de vocabulário é usado para melhorar a eficácia dos sistemas de armazenamento e recuperação da informação, sistemas de navegação na *Web* e outros ambientes que visam identificar e localizar conteúdos desejados através da descrição, usando algum tipo de linguagem.” (National Information Standards Organization, 2010, p. 1, tradução nossa)⁵. Quanto à poli-hierarquização, de acordo com Cavalcante e Bräscher (2014), esta pode ser mais facilmente encontrada nas taxonomias navegacionais: “Deve ser empregada sempre que existir mais de uma possibilidade lógica de uma categoria, subcategoria ou elemento que possa ser encontrado dentro da estrutura.” (Cavalcante; Bräscher, 2014, p. 200).

A forma de organização da estrutura taxonômica depende da característica selecionada como critério para subdivisão. Esta escolha do critério não é neutra, mas sim, delimitada pelas características da organização que utilizará a taxonomia e pela finalidade para a qual a taxonomia foi construída (Schiessl; Shintaku, 2012): “Não há taxonomia certa ou errada, nem geral o suficiente para abranger todos os aspectos da realidade.” (Schiessl; Shintaku, 2012, p. 85).

Schiessl e Shintaku (2012) destacam o aspecto dinâmico das taxonomias, que são mais facilmente adaptáveis às alterações do ambiente que representam e do conhecimento gerado por quem as utiliza. Esta característica desperta a atenção no contexto empresarial (Schiessl; Shintaku, 2012), devido à necessidade de gerir o conhecimento para fins estratégicos e de competição por mercado (Aganette; Alvarenga; Souza, 2010).

Atualmente, as taxonomias estão sendo utilizadas como instrumentos para a organização e recuperação da informação em ambientes organizacionais e apoio a gestão. Sua aplicação, pode ser direcionada a portais na *web*, tendo como característica, o uso dos recursos inerentes a esse ambiente (Vital; Café, 2007). De acordo com Vital e Café (2007), as taxonomias corporativas não requerem aplicação de critérios rígidos na construção de suas relações semânticas. No ambiente organizacional, o principal objetivo é a significação aos grupos de usuários para os quais a taxonomia foi desenvolvida (Vital; Café, 2007).

⁵ No original: “Vocabulary control is used to improve the effectiveness of information storage and retrieval systems, Web navigation systems, and other environments that seek to both identify and locate desired content via some sort of description using language.” (National Information Standards Organization, 2010, p. 1).

As taxonomias corporativas caracterizam-se por (Campos; Gomes, 2007, p. 3-4):

- Conter uma lista estruturada de conceitos/termos de um domínio;
- Incluir termos sem definição, somente com relações hierárquicas;
- Possibilitar a organização e recuperação de informação através de navegação;
- Permitir agregação de dados, diferentemente das taxonomias seminais, além de evidenciar um modelo conceitual do domínio;
- Ser um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em um Sistema de Recuperação de Informação;
- Ser um novo mecanismo de consulta em Portais institucionais, através de navegação.

As taxonomias caracterizam-se como SOCs, apresentando os seguintes aspectos inerentes (Cavalcante; Bräscher, 2014). O primeiro se refere ao princípio da categorização, que define a taxonomia como instrumento destinado à classificação de elementos a partir de características compartilhadas. Esta é uma acepção tradicional da taxonomia, mas este aspecto pode ser observado atualmente na aplicação de taxonomias em páginas na *web*. O segundo aspecto refere-se às relações semânticas. Apesar da estrutura taxonômica apresentar essencialmente relações hierárquicas, ela é flexível, podendo aceitar a ocorrência de outros tipos de relações semânticas, como, por exemplo, a existência de taxonomias facetadas. O terceiro aspecto se refere a função auxiliar da taxonomia para navegação em páginas na *web*, orientando a navegação do usuário e organizando as informações e *links* de forma lógica, por meio das categorias dispostas em estruturas hierárquicas (Cavalcante; Bräscher, 2014).

A utilização de taxonomias para navegação na *web* é considerada por Zhonghong *et al.* (2006) como o principal papel desempenhado pelas taxonomias atualmente. Neste sentido, os autores identificam dois elementos-chaves destes instrumentos: a estrutura hierárquica, servindo para a organização da informação na página na *web*, e os rótulos, representando os conceitos compreendidos nesta estrutura (Cavalcante; Bräscher, 2014).

Para Martinez *et al.* (2004, p. 107, tradução nossa):

[...] a taxonomia, em um sentido amplo, é a criação da estrutura (ordem) e dos rótulos (nomes) que ajudam a localizar a informação relevante. Em um

sentido mais específico, é o ordenamento e rotulação de metadados, que permite organizar sistematicamente a informação primária.⁶

A taxonomia, portanto, serve como instrumento para a organização da informação no ambiente digital, tanto da perspectiva da navegação em páginas na *web*, quanto como rotulação de metadados, auxiliando na indexação de conteúdos na *web*, possibilitando a recuperação da informação por motores de busca (Schiesl; Shintaku, 2012).

As taxonomias vêm sendo usadas para a criação de metadados ou termos comuns para descrever um objeto, com foco na recuperação da informação e na categorização, como suporte de navegação e esquemas que organizam conteúdos das páginas na *web* e lista de controle de dados usados para suporte de mineração de dados. (Carlan; Medeiros, 2011, p. 57).

As taxonomias representam os objetivos de organização do conhecimento de um contexto específico. Portanto, elas podem se diferenciar, dependendo do tipo de organização e informação que visam representar. Campos e Gomes (2007) compreendem a definição de três tipos de taxonomias: canônica (classificação binária, dicotômica), de domínio e de processos e tarefas. As taxonomias de domínio e de processos são policotômicas. Como várias subdivisões são possíveis, no caso destes dois tipos de taxonomia, é importante primeiramente definir o recorte do domínio que a taxonomia pretende representar. A taxonomia com princípio policotômico de classificação, possibilita a organização a partir de um critério de divisão, que é aplicado em todas as suas subdivisões (Campos; Gomes, 2007).

No contexto organizacional, as taxonomias podem ser divididas em: taxonomia descritiva, de navegação ou para gerenciamento de dados (Conway; Sligar, 2002). A taxonomia descritiva consiste em um vocabulário controlado, com estabelecimento de termos relevantes em determinado contexto, apresentando relações de sinonímia e homonímia (Vital, 2007), possibilitando a categorização do conteúdo informacional produzido pela organização, considerando sua terminologia (Cavalcante; Bräscher, 2014). Devido a estas características, assemelha-se ao modelo de um tesouro, em que os termos preferidos são relacionados aos seus equivalentes preteridos, de maneira que a busca por qualquer forma do termo recupere a informação pesquisada (Cavalcante; Bräscher, 2014).

⁶ No original: “[...] la taxonomía es, en un sentido amplio, la creación de la estructura (orden) y de los rótulos (nombres) que ayudan a localizar la información relevante. En un sentido más específico, es el ordenamiento y rotulación de los metadatos, que permiten gestionar sistemáticamente la información primaria.” (Martinez *et al.*, 2004, p. 107).

A taxonomia de navegação possibilita a recuperação da informação por meio da navegação através de suas categorias, sendo amplamente adotada na organização da informação em páginas na *web* (Cavalcante; Bräscher, 2014). A estrutura da taxonomia fica visível ao usuário final e, com base na observação do comportamento de navegação do mesmo, a taxonomia pode ir sendo modificada (Vital; Café, 2007). Vital (2007, p. 48) declara que: “As relações entre os termos precisam fazer sentido para os usuários dessa taxonomia, sem necessariamente apresentarem as subordinações lógicas.”

Cavalcante e Bräscher (2014) destacam que, relacionadas às taxonomias navegacionais, existem as taxonomias navegacionais facetadas (Maculan; Lima, 2011), também chamadas de multidimensionais (Aganette; Alvarenga; Souza, 2010), em que uma categoria é composta por diversos aspectos (facetas) que constituem o seu objeto. Por exemplo, em um *site* de comércio eletrônico, a categoria “Cafeteira” poderia apresentar as facetas “material”, “voltagem” e “potência”. De acordo com Cavalcante e Bräscher (2014), as facetas podem ser empregadas como filtros que ajudam a refinar categorias extensas. Este tipo de taxonomia está encontrando espaço em produtos comerciais, que visam a organização da informação no comércio eletrônico ou em ambientes corporativos (Aganette; Alvarenga; Souza, 2010).

A taxonomia para gerenciamento de dados contém um pequeno conjunto de termos controlados, referentes a um grupo específico da organização, como um setor ou departamento, porém sem qualquer estrutura hierárquica entre os termos (Aganette; Alvarenga; Souza, 2010; Vital, 2007).

Cavalcante e Bräscher (2014) sumarizam as principais diferenças entre as taxonomias descritivas e navegacionais aplicadas em ambientes digitais:

Em síntese, tem-se que os dois tipos definidos de taxonomia atuam no apoio à recuperação da informação de formas distintas: enquanto a taxonomia descritiva serve de suporte aos mecanismos de busca (*searching*), as taxonomias navegacionais estão relacionadas à recuperação da informação a partir da navegação por seus descritores (*browsing*). Por outro lado, acrescenta-se que as taxonomias descritivas são transparentes aos usuários, ao passo que as taxonomias navegacionais se inserem dentre os elementos gráficos utilizados para a navegação *Web*. (Cavalcante; Bräscher, 2014, p. 194).

Para a elaboração de uma taxonomia, é necessário seguir alguns critérios gerais, elencados por Campos e Gomes (2007, p. 9) conforme abaixo:

- Comunicabilidade: o termo empregado deve representar a linguagem utilizada pelo usuário;

- Utilidade: o nível de especificidade dos termos deve expressar um agrupamento de documentos e não um único documento, ou seja, o termo só é útil quando representativo para um conjunto de documentos;
- Estimulação: uso de termos que induzem o usuário a continuar a navegação pelo sistema;
- Compatibilidade: os termos empregados representam o campo que se está ordenado, fazendo parte das atividades e funções da organização.

Campos e Gomes (2007) sistematizam proposta para a elaboração de taxonomia, dividindo em quatro etapas. A primeira etapa refere-se à captura do conhecimento, em que é realizado o levantamento dos assuntos que farão parte da estrutura da taxonomia. Este levantamento pode ser realizado por meio de entrevistas com especialistas, documentos da instituição ou outros instrumentos classificatórios ou terminológicos (Campos; Gomes, 2007).

A segunda etapa compreende a análise dos documentos e objetos informacionais que serão agregados à taxonomia. Após, a terceira etapa compreende a elaboração da estrutura classificatória. Nesta etapa, primeiramente, é importante definir a tipologia de taxonomia que será adotada. Também, deve-se definir a forma pela qual a taxonomia será representada e qual *software* será utilizado para esta representação, pois, muitas vezes, dependendo do *software* selecionado, pode-se restringir as possibilidades de representação da taxonomia (Campos; Gomes, 2007).

Vital e Café (2007), a partir de estudos sobre as práticas de elaboração de taxonomia, compilam as etapas para a construção da sua estrutura classificatória:

- a) estabelecimento das categorias gerais: deve seguir uma abordagem dedutiva e indutiva. Segundo Vickery (1960), no método dedutivo são estabelecidas as categorias gerais por meio de um trabalho envolvendo especialistas, profissionais da informação e usuário. Já o método indutivo busca definir as categorias gerais com base nas principais fontes de informação na área;
- b) coleta de termos: compreende a consulta à literatura especializada ou instrumentos classificatórios já existentes na área, como um recurso auxiliar. Os termos devem fazer sentido para o usuário e a literatura serve como validação dos termos selecionados, atingindo, então, a garantia literária e a garantia de uso;
- c) análise dos termos selecionados: compreende a padronização da terminologia empregada e a análise das variações terminológicas;

- d) controle da diversidade de significação: compreende o controle de ambiguidades por meio da contextualização do significado do termo, através da elaboração de definições e notas de escopo;
- e) construção dos relacionamentos semânticos: considera as relações semânticas de equivalência, hierárquicas e não-hierárquicas, bem como a polihierarquia.

Seguindo com as quatro etapas propostas por Campos e Gomes (2007) para elaboração de taxonomia, a última etapa compreende a validação, em que a estrutura classificatória construída é validada por especialistas, com objetivo de confirmar o trabalho realizado e transferir o conhecimento do processo de elaboração. Campos e Gomes (2007) destacam a importância de se realizar a validação em cada uma das etapas no processo de elaboração de taxonomia, a fim de garantir o atingimento dos critérios de comunicabilidade, estimulação e compatibilidade (Campos; Gomes, 2007).

Por fim, Campos e Gomes (2007, p. 10-11) consideram que:

O sucesso dos empreendimentos se deve, em especial, a dois aspectos a serem considerados: primeiro, a adoção de princípios teóricos e metodológicos de classificação consistentes, para que as taxonomias possam ser elaboradas de forma a serem expandidas e, segundo, estar atento para soluções relativas à arquitetura de informação existente para acomodar as taxonomias.

Concluindo, a taxonomia mostra-se como um instrumento para a organização e representação do conhecimento e da informação, especialmente em ambientes organizacionais e digitais, como em *sites* de comércio eletrônico, contexto desta pesquisa. Cavalcante e Bräscher (2014) consideram que este tipo de ambiente digital figura entre os principais nichos de utilização de taxonomias, especialmente as taxonomias navegacionais, que podem atuar com papel preponderante na decisão de compra de determinado produto ou serviço. A organização e representação dos produtos em um comércio eletrônico, por meio da elaboração e adoção de uma taxonomia, busca melhorar a experiência de compra do usuário, seja por meio da busca ou através da navegação no ambiente digital (Steimer; Luz, 2015).

2.1.1.3 Tesouro

O tesouro é um vocabulário controlado, podendo ser considerado atualmente como um tipo de SOC (Hjørland, 2016a). É formado por conceitos representados por

termos relacionados entre si através de relações hierárquicas, associativas e de equivalência, constituindo uma representação da estrutura conceitual lógica e semântica de um domínio do conhecimento (Schiessl; Shintaku, 2012; Campos, 2021; Gabriel Junior; Laipelt, 2019).

No contexto da CI, o tesouro atua como um instrumento de controle terminológico para traduzir a linguagem natural dos documentos e dos usuários (pesquisadores e indexadores) para um sistema, apresentando termos de forma unívoca, neutralizando ambiguidades, visando garantir a consistência no tratamento do conteúdo dos documentos (Campos, 2021).

De acordo com Schiessl e Shintaku (2012, p. 96):

A característica da linguagem documentária é a tradução da linguagem natural para uma linguagem padronizada e estável. No ambiente da ciência da informação, a busca pela melhoria e efetividade dos sistemas de gestão da informação é um objetivo primordial. Para isso, é fundamental alcançar consistência na descrição dos documentos que estão armazenados nos sistemas de informação. Esse é o objetivo primário no controle de vocabulário.

A necessidade de se controlar o vocabulário origina-se nas duas características básicas da língua natural (Schiessl; Shintaku, 2012):

- a) duas ou mais palavras podem ser utilizadas para representar o mesmo conceito;
- b) duas ou mais palavras com a mesma escrita podem representar conceitos diferentes.

Segundo Schiessl e Shintaku (2012), existe consenso sobre as duas principais funções do tesouro, a saber:

- a) representação dos assuntos de um documento: o indexador, no momento da classificação, analisa o documento que quer representar, identifica os termos que melhor se relacionam ao assunto, e os exprime na linguagem específica do tesouro. A redução da dispersão dos elementos relacionados se consegue com o controle (agrupamento) dos sinônimos e quase sinônimos, com a distinção dos homógrafos e com a reunião dos termos que possuem uma relação mais próxima entre si (Boccatto; Ramalho; Fujita, 2008);
- b) auxílio na recuperação da informação: auxilia na formulação de consultas feitas pelos usuários, facilitando a pesquisa, atrelando os termos com relações sintagmáticas ou paradigmáticas (Boccatto;

Ramalho; Fujita, 2008). Com o apoio do tesouro, o usuário consegue expandir ou restringir sua busca, além de dar ao usuário outras opções de busca em relação ao assunto procurado (Dodebei, 2002).

De acordo com Carlan e Medeiros (2011), a estrutura dos tesouros ocorre em duas partes: primeiro a base teórica, que é a essência e a fundamentação para a construção de tesouros. São o conceito, o termo, as categorias e a faceta. Já a base técnico-operacional é a abordagem mais prática no desenvolvimento de tesouros, contemplando o planejamento, a coleta de termos, o controle terminológico, o estabelecimento de relações entre conceitos e as formas de divulgação e publicação dos tesouros.

Os principais elementos do tesouro são os termos, representando os conceitos, e as relações entre eles, estabelecendo uma estrutura semântica. Segundo Dahlberg (1978a), um conceito é uma unidade de conhecimento que consiste em afirmações verdadeiras sobre um referente expresso em palavras. Cada afirmação sobre um referente (objeto, ação) é composta por um elemento do respectivo conceito, identificado como característica (propriedade) desse conceito. Os conceitos são compostos por várias características, e características semelhantes levam a relacionamentos entre conceitos. Esses relacionamentos podem ser analisados e interpretados para permitir a organização em categorias e subcategorias.

A noção de conceito como unidade estrutural no tesouro surge com a proposta da classificação facetada de Ranganathan, culminando no campo de estudos atualmente denominado Organização do Conhecimento, cujas bases teóricas fundamentam-se em teorias da terminologia e na Teoria do Conceito de Dahlberg (Dodebei, 2002). A ideia de conceito e não mais de assunto para representar o conhecimento é caracterizada por uma estrutura pós-coordenada, ou seja, estrutura que se forma por relações conceituais e não por relações de assunto. Exemplos desta estrutura conceitual são os tesouros (Dodebei, 2002).

O conceito é representado por um termo, que por sua vez é uma palavra especializada, ou conjunto delas (Schiessl; Shintaku, 2012). É a menor unidade de representação em um tesouro. Guarda relação intrínseca com o referente, no campo conceitual (Campos, 2021). Ao contrário da palavra, no léxico comum, o termo possui apenas um significado, porque se relaciona univocamente com o conceito especializado em uma área do conhecimento (Schiessl; Shintaku, 2012).

Os termos, como unidades de representação do conceito, se relacionam entre si, formando um sistema conceitual:

Para o estabelecimento de comparações por semelhanças e diferenças entre os conceitos, quer seja para defini-los, nomeá-los ou ordená-los, há que se falar com relações conceituais. Do ponto de vista epistemológico, o conceito pode ser analisado segundo as ações mentais, ou seja, a habilidade da mente em comparar algo novo com conhecimento adquirido. (Dodebei, 2002, p. 90-91).

As relações entre os termos no tesauro são de vários tipos e podem ser designadas como relação hierárquica, relação partitiva, relação associativa, relação de equivalência e relação de instância (Campos, 2021):

- a) relação hierárquica (hiperonímia e hiponímia): gênero/espécie. Quando dois conceitos diferentes possuem características idênticas e um deles possui uma característica a mais do que o outro (Dodebei, 2002). Indicadores de relacionamento no tesauro: TG (termo genérico) / TE (termo específico);
- b) relação partitiva (meronímia): partes de um objeto, elementos de um sistema, etapas de um processo ou atividade. Indicadores de relacionamento no tesauro: TGP (termo genérico partitivo) / TEP (termo específico partitivo);
- c) relação associativa: no tempo e no espaço, entre conceitos fora da mesma hierarquia. Indicadores de relacionamento no tesauro: TA (termo associado) ou TR (termo relacionado);
- d) relação de equivalência: ocorrem entre dois objetos com as mesmas características e que possuem formas de designação diferentes, neste caso, os termos são considerados sinônimos. Correspondem à relação entre o termo preferido (descriptor) e o não-preferido (não-descriptor) (Bocato; Ramalho; Fujita, 2008). Segundo Dodebei (2002), funcionam para evitar a dispersão léxica, simbólica e sintática. Indicadores de relacionamento no tesauro: USE (relação usada no termo não-preferido para o termo preferido) e UP (relação usada no termo preferido para o não preferido);
- e) relação de instância: relação entre um objeto geral e um objeto individual (nomes próprios, instâncias). Indicadores de relacionamento no tesauro: TGI (termo genérico de instância) e TEI (termo específico de instância).

Quando um mesmo símbolo é identificado com dois ou mais conjuntos de características distintas, é configurada a polissemia. Neste caso, os símbolos devem ser diferenciados através de qualificadores, o que os transformará em novos símbolos, cada um no lugar de um único significado conceitual (Dodebei, 2002).

As relações conceituais podem ser divididas entre relações paradigmáticas e sintagmáticas. As relações hierárquicas, partitivas e de equivalência constituem as relações paradigmáticas, pois ocorrem dentro da mesma categoria. Já as relações associativas constituem as relações sintagmáticas, pois ocorrem entre entidades de diferente natureza (Gabriel Junior; Laipelt, 2019). As relações conceituais serão detalhadas mais adiante neste trabalho, na seção 2.1.2.2 Teoria do Conceito.

Considerando o tesauro como uma estrutura conceitual que representa um domínio do conhecimento, além de um instrumento de controle terminológico, é importante que o aspecto semântico desta estrutura seja enriquecido. Isso se dá através das relações sintagmáticas, que expandem as possibilidades semânticas das relações contidas no tesauro. Para isso, é necessária a explicitação das relações associativas, formalizando-as. Há trabalhos neste sentido, como em Gabriel Junior e Laipelt (2019), que analisam como as relações associativas podem ser explicitadas em tesouros, construindo conexões semânticas mais fortes entre os conceitos, como ocorre em ontologias.

A estrutura do tesauro baseia-se de maneira lógico-semântica. Apresenta uma parte sistemática e uma parte alfabética. Na parte sistemática, o conjunto de conceitos se apresenta na vertical (relações hierárquicas, formando relações lógicas) e na horizontal (relações não-hierárquicas, associando-se semanticamente), como uma taxonomia, que permite ter uma visão abrangente do domínio representado. Na parte alfabética, cada conceito é apresentado com seus indicadores de relações, notas de definição e notas de escopo (ou notas de aplicação) (Bocato; Ramalho; Fujita, 2008; Campos, 2021).

Para a elaboração de um tesauro, é importante iniciar pelo recorte classificatório, ou seja, decidir qual recorte do domínio específico será representado. A definição do domínio e recorte está relacionada à Análise de Domínio e será detalhada na seção 3.4 ANÁLISE DE DOMÍNIO. De acordo com Campos (2021), os recortes podem ser assim divididos:

- a) recorte por disciplina: “[...] o domínio é classificado por áreas canônicas, ou seja, por disciplinas e subdisciplinas.” (Campos, 2021, p. 90);

- b) recorte pela extensão das categorias de um campo de assunto: “[...] é realizado a partir da determinação de categorias selecionadas deste campo, no interior do qual se determina facetas e seus aspectos relacionados.” (Campos, 2021, p. 90). Categorias são conceitos abrangentes acima das quais não se pode colocar qualquer outro conceito;
- c) recorte pela estrutura organizacional: “[...] o domínio é classificado por setores ou funções de uma organização.” (Campos, 2021, p. 90).

Dentro do recorte, os termos são organizados por classes de conceitos, com base em características de divisão. Estas características são os princípios utilizados para a formação de classes que são formadas por cadeias e renques (Campos, 2021):

- a) cadeia: é uma série vertical de conceitos, em que cada conceito tem uma característica a mais ou a menos conforme a cadeia seja descendente ou ascendente;
- b) renque: são as classes formadas a partir de uma única característica de divisão, formando séries horizontais. São os termos coordenados entre si.

Campos (2021, p. 91) destaca que: “Ranganathan (1967)⁷ apresenta uma série de princípios (cânones) para garantir a consistência na formação de renques e cadeias (cânones para sucessão e características, cânones para a formação de renques e cânones para a formação de cadeias).” Estes princípios serão explicados mais detalhadamente na seção 2.1.2.1 Teoria da Classificação Facetada.

Quanto a elaboração e manutenção de tesouros, Boccato, Ramalho e Fujita (2008, p. 201) apresentam proposta de metodologia, conforme abaixo:

- a) planejamento: delimitação do domínio; classificação (método de categorização); levantamento das fontes de informação; forma de apresentação (alfabética/sistemática); período de atualização; seleção do software de construção; divulgação e manutenção do tesouro;
- b) levantamento dos termos: escolha da forma de apresentação (termos simples e/ou compostos);
- c) organização dos termos/conceitos (aplicação de fichas terminológicas);
- d) apresentação final (alfabética/sistemática): definição da forma gráfica aplicada a ordem sistemática; impresso e/ou eletrônico;
- e) interoperabilidade: definição de necessidades para a ocorrência da interoperabilidade entre o sistema e o tesouro (referentes ao processo de indexação e busca da informação);
- f) definição de critérios para testes e avaliação do tesouro.

⁷ RANGANATHAN, S. R. **Prolegomena to Library Classification**. 3. ed. New York: Asia Publishing House, 1967.

A seleção dos termos configura-se na extração de palavras e frases da literatura do domínio de conhecimento específico. A coleta de termos na literatura especializada contribui na identificação de termos monossômicos, pois foram utilizados por especialistas no domínio em que estão inseridos (Schiessl; Shintaku, 2012).

A ANSI/NISO Z39.19:2005 orienta sobre os princípios de validação, que pode se dar por meio da garantia literária e da garantia de usuário. A garantia literária sustenta que a literatura determina o universo de termos que serão utilizados para representar o domínio pretendido. Já a garantia de usuário é a confirmação de que os termos do tesouro estão de acordo com quem usa o sistema, ou seja, o usuário (Schiessl; Shintaku, 2012). Elas não são concorrentes, mas sim, complementares e importantes da mesma forma.

2.1.1.4 Ontologia

A ontologia tem sua origem na filosofia, sendo um ramo da metafísica. Ela estuda a natureza das coisas e define categorias fundamentais. Na área da computação, mais especificamente no campo de engenharia do conhecimento e inteligência artificial, a ontologia é compreendida como um artefato ou esquema conceitual que define conceitos, relacionamentos entre eles e suas propriedades, com a finalidade de representar um domínio de conhecimento, por meio de uma linguagem inequívoca e legível por máquina (Guarino; Oberle; Staab; 2009).

Uma definição bastante comum de ontologia encontrada na literatura é a de uma especificação explícita e formal de uma conceitualização compartilhada em um domínio de conhecimento (Gruber, 1995), em que especificação explícita refere-se a conceitos, atributos, relações e axiomas explicitamente definidos. Nesse contexto, formal significa legível por máquinas (computadores); conceitualização pode ser entendida como um modelo abstrato de algum fenômeno do mundo real; e compartilhada significa conhecimento consensual (Mendonça; Zaidan, 2019).

Ou seja, a ontologia em um sistema de informação é concebida como um artefato que representa uma estrutura conceitual, a partir de uma conceitualização compartilhada e consensual em um domínio do conhecimento, por meio de uma

linguagem formal e explícita que pode ser entendida por computadores. O objetivo é que a ontologia provenha às máquinas uma estrutura de mundo (Almeida, 2020).

Ontologia abrange uma estrutura classificatória usada comumente em taxonomias e tesouros. Porém, sua principal característica é a apresentação de propriedades de cada classe, além das relações hierárquicas e associativas. Desta forma, a ontologia funciona como um vocabulário conceitual e um modelo de trabalho que permite armazenar, pesquisar e raciocinar com base em instâncias e regras (Zeng, 2008). Dahlberg (1996, p. 129) destaca este aspecto, declarando que: “[...] ontologia tem de fato algo a ver com sistemas de classificação no sentido de que o que precisamos organizar são nossos conceitos sobre a realidade [...]”.

Uma das principais aplicações da ontologia é na *web* semântica, chamada também de *web* 3.0 ou *web* de dados, da qual é uma das bases (W3C Brasil, 2023). De acordo com W3C Brasil (2023): “O objetivo final da *Web* de dados é possibilitar com que computadores façam coisas mais úteis e com que o desenvolvimento de sistemas possa oferecer suporte a interações na rede.” Na *web* semântica, a *internet* não está mais baseada em documentos, mas em dados conectados. A ontologia serve como um vocabulário, uma linguagem comum destes dados, por meio de seus metadados, para que sejam representados através de uma estrutura conceitual, que dê sentido aos termos contidos nestes dados.

A ontologia é um artefato representacional, definido como uma “[...] representação codificada por regras interpretativas - públicas, claras e precisas - que permitem entender como o artefato se refere à realidade.” (Almeida, 2020, p. 44). Para se criar um artefato representacional, são adotados vários constructos. Abaixo são listados os principais componentes de uma ontologia (Biagetti, 2021; Mendonça; Zaidan, 2019; Madsen; Thomsen, 2015; Almeida, 2020; International Organization for Standardization, 2011):

- a) entidades: trata-se de um dos termos mais genéricos para referenciar algo, próximo ao termo “coisa”;
- b) classes: reúnem entidades de um domínio, com uma certa similaridade;
- c) hierarquia de classes: formada através de subclasses entre as classes. Todas as características de uma classe superordenada se aplicam às suas respectivas subclasses;
- d) instâncias: exemplos ou casos particulares das classes; indivíduos usados para representar objetos e conceitos específicos;

- e) relacionamentos: representam o tipo de interação entre as classes do domínio, estabelecem conexões entre classes e outras classes fora da mesma hierarquia;
- f) propriedades: caracterizam as instâncias que compõem uma classe. Elas utilizam restrições de domínio (*domain*) e abrangência (*range*), definindo com precisão as classes de indivíduos que podem ser conectados pela propriedade, além de definir a quantidade máxima e mínima de indivíduos ou valores de atributos que podem estar conectados (cardinalidade). As propriedades podem ser subdivididas em (Biagetti, 2021):
 - propriedades de objeto: conectam um indivíduo a outro indivíduo, definindo como os indivíduos são relacionados, e;
 - propriedades de dados: são usadas para atribuir valores de dados a objetos;
- g) axiomas: declarações e definições que são sempre consideradas verdadeiras, como as definições de classes; são regras formais usadas para restringir a interpretação e o uso dos termos envolvidos, geralmente especificadas em uma linguagem formal de lógica;
- h) rótulos e identificadores: é o termo em linguagem natural conferido a uma classe, propriedade ou indivíduo, para sua identificação por humanos.

A estrutura base da ontologia é uma taxonomia, pois nela se estabelecem as relações hierárquicas entre classes e subclasses, utilizando relacionamento é-um. A partir da estrutura taxonômica, estabelecem-se relacionamentos entre classes de diferentes hierarquias e são trazidos atributos para definir estas classes. Conforme Almeida (2020), quando se agrega outros tipos de relacionamentos que não sejam taxonômicos a uma taxonomia, obtém-se uma ontologia.

Quando são instanciadas entidades individuais nas classes da ontologia e lhes são atribuídos valores de propriedades, é formada uma base de conhecimento (Noy; McGuinness, 2001). A partir desta base de conhecimento, é possível fazer inferências, descobrir novos conhecimentos, recuperar informação, além da aplicação em sistemas de informação e algoritmos de inteligência artificial.

Existem algumas metodologias para o desenvolvimento de ontologias, dentre elas, a metodologia 101, que é dividida em sete etapas, sendo elas (Noy; McGuinness, 2001):

- a) determinar o domínio e escopo da ontologia: entender a abrangência, finalidade, pessoas envolvidas e o que a ontologia deve responder (questões de competência);
- b) verificar possibilidade de reutilizar ontologia já existente: neste caso, há métodos específicos para a interoperabilidade entre ontologias;
- c) listar termos que possam ser importantes para a ontologia;
- d) definir as classes e a hierarquia destas classes:
 - após o passo anterior, seleciona-se os termos que descrevem objetos com existência independente, estas serão as classes;
 - então, organizam-se estas classes em uma hierarquia a partir de alguma destas abordagens:
 - de baixo para cima (*bottom-up*): definição das classes mais específicas e agrupamento em classes mais genéricas;
 - de cima para baixo (*top-down*): definição das classes mais genéricas e subsequente especialização dos conceitos;
 - uma combinação das duas formas: mais indicada para iniciantes, selecionam-se os termos mais destacados para então generalizá-los e especializá-los;
- e) definir as propriedades das classes (*slots*): após a etapa anterior, possivelmente os termos que sobraem da lista irão configurar-se como os atributos das classes. Para cada propriedade da lista, determina-se qual classe ela descreve. Essas propriedades tornam-se *slots* anexados às classes. O *slot* deve ser anexado a classe mais genérica que tenha estas propriedades, e todas as subclasses herdarão o *slot* da classe superordenada. Existem alguns tipos de propriedades que podem se tornar *slots* na ontologia:
 - propriedade intrínseca: atributos que fazem parte da entidade;
 - propriedade extrínseca: atributos que são atribuídos à entidade;
 - partes da entidade, podendo ser físicas ou abstratas;
 - relacionamentos entre instâncias;

- f) definir as facetas dos *slots*: as facetas do *slot* descrevem os valores permitidos do atributo, o tipo deste valor, o número dos valores (cardinalidade) e outras características dos valores que o *slot* pode assumir:
- cardinalidade: quantos valores o *slot* pode salvar;
 - tipo de valor: pode ser texto, numérico, booleano, enumerativo e instâncias de uma classe;
 - domínio (*domain*) e abrangência (*range*): domínio refere-se às classes às quais um *slot* está anexado ou classes cuja propriedade um *slot* descreve, já a abrangência refere-se às classes permitidas para *slots* do tipo Instância;
- g) criar instâncias: até então, o desenvolvimento estava na fase de modelagem. Após a definição e organização de toda a estrutura da ontologia, instanciam-se indivíduos reais no modelo desenvolvido, gerando uma base de conhecimento.

Existem sistemas específicos para a construção de ontologias, como o Protégé, um editor de ontologia de código aberto para a construção de sistemas inteligentes, desenvolvido pela Universidade de Stanford, nos Estados Unidos (Protégé, 2023). A partir da modelagem conceitual em editores de ontologia, é possível transcrever esta estrutura para uma linguagem de ontologia que seja legível por máquinas e também que seja própria para *web*. “Uma linguagem de ontologia deve descrever significados formalmente e de forma legível por máquina para permitir o raciocínio automatizado.” (Biagetti, 2021, p. 167, tradução nossa).⁸

Os formatos mais utilizados para a representação formal de ontologias são as linguagens *Resource Description Framework* (RDF) e *Ontology Web Language* (OWL), ambas criadas e padronizadas pelo W3C (Mendonça; Zaidan, 2019). A partir de uma ontologia representada por meio de uma linguagem formal, algoritmos podem ser usados para produzir inferências, determinando se um indivíduo pertence a uma classe, estabelecendo relacionamentos hierárquicos entre conceitos e analisando a consistência lógica entre conceitos (Biagetti, 2021).

As ontologias podem ser classificadas de diferentes formas, de acordo com a abrangência do escopo, o nível de detalhamento do domínio e finalidade da aplicação

⁸ No original: “An ontology language should describe meanings formally and in a machine-readable way to allow for automated reasoning.” (Biagetti, 2021, p. 167).

(International Organization for Standardization, 2011). Elas podem ser divididas em ontologias de domínio, ontologias de referência ou ontologias de alto nível (Farinelli; Souza, 2021).

As ontologias de alto nível (*top-level*) representam conceitos e propriedades universais, que independem de um domínio do conhecimento específico. A principal função de ontologias de alto nível é permitir a cooperação entre múltiplos agentes artificiais e humanos (Biagetti, 2021) e a interoperabilidade entre as ontologias de domínio (Farinelli; Souza, 2021). As ontologias de domínio descrevem as classes de um domínio específico do conhecimento, podendo ser baseadas em ontologias de alto nível, especificando as classes gerais fornecidas por estas (International Organization for Standardization, 2011; Biagetti, 2021).

Observa-se que as ontologias, dentre os exemplos de SOCs detalhados nas seções anteriores, se mostra como uma opção mais adequada para sistemas de informação, visto que adota uma linguagem formal entendida por máquinas e se propõe a ser uma linguagem comum para a estruturação e interoperabilidade de dados no contexto da web semântica. Um aspecto que não é muito tratado no âmbito da ontologia é o controle terminológico, em contraponto com o tesouro, que apresenta principalmente esta finalidade. Portanto, identifica-se oportunidade de explorar as qualidades de cada modelo, buscando o benefício que a implementação de cada um deles pode trazer para determinado contexto.

2.1.2 Bases teóricas para desenvolvimento de Sistemas de Organização do Conhecimento

Hjørland (2016b) considera menos importante a discussão sobre formas específicas de SOCs. Acredita ser mais importante fazer uma abstração para sistemas de conceitos e suas relações semânticas e para entender cada tipo específico de SOC como sistema baseado em princípios que são gerais para todos os SOCs.

Portanto, é importante que seja observada base teórica para o desenvolvimento de SOCs, que pode ser encontrada na Teoria da Classificação, na Teoria do Conceito e nas teorias da Terminologia para o tratamento dos conceitos e termos em estruturas como SOCs (Carlan; Medeiros, 2011).

2.1.2.1 Teoria da Classificação Facetada

A classificação é um processo inerentemente humano, transformando impressões isoladas em conceitos reconhecíveis, o que ajuda a entender o mundo (Schiessl; Shintaku, 2012). “O ato de classificar é intrínseco aos processos mentais humanos, quando separamos ‘coisas’ em grupos distintos, com características comuns entre si, estamos classificando.” (Pinho; Vital, 2016, p. 97). A classificação pode ser definida como a “[...] formação metódica e sistemática de grupos onde se estabelecem critérios para a divisão.” (Carlan; Medeiros, 2011, p. 59). Os critérios para agrupamento e divisão são diversos, pois de acordo com o contexto e a finalidade da classificação, os objetos e conceitos podem ser organizados de diferentes perspectivas. Portanto, não há uma classificação única, mas uma classificação de acordo com o contexto e a finalidade da classificação (Schiessl; Shintaku, 2012).

A classificação está diretamente relacionada com a organização do conhecimento, visto que é uma forma pela qual é possível organizar e representar uma determinada área do conhecimento. A classificação possui dois processos principais: a elaboração de esquemas de classificação e o seu uso (Schiessl; Shintaku, 2012). Na perspectiva da CI, os esquemas de classificação surgiram como uma tentativa de resposta às demandas práticas na gestão de materiais bibliográficos em bibliotecas, tendo a pretensão de cobrir a totalidade do conhecimento humano (Pinho; Vital, 2016).

Dentre as teorias da classificação, destaca-se a Teoria da Classificação Facetada, também conhecida como Teoria Analítico-Sintética, desenvolvida por Shiyali Ramamrita Ranganathan, matemático e bibliotecário indiano, grande teórico da Classificação no âmbito da CI (Gomes; Campos, 2019). Ranganathan contestou os sistemas de classificação que eram comumente adotados em bibliotecas, como a Classificação Decimal de Dewey (CDD), visto que nestes sistemas, as classes precisam ser definidas antes dos objetos serem classificados e também não se mostravam adequados para a representação de assuntos compostos, pois, por se tratarem de uma estrutura hierárquica, não possibilitam a representação de mais de um aspecto do objeto ou conceito ao mesmo tempo (Pontes; Lima, 2012).

Ao invés de criar compartimentos para depois colocar os objetos dentro dos mesmos, Ranganathan buscava uma forma de partir do objeto e, então, coletar e arranjar aspectos (ou facetas) que pudessem descrever aquele

objeto. Essa abordagem possibilita maior flexibilidade e um alto grau de especificidade no sistema de classificação. (Pontes; Lima, 2012, p. 23).

A partir da década de 1920, Ranganathan iniciou o desenvolvimento de sua teoria e de um sistema de classificação facetada. Pela primeira vez, era desenvolvida uma teoria que embasava um sistema de classificação (Pinho; Vital, 2016), fornecendo uma metodologia racional e lógica para o desenvolvimento de esquemas de classificação, ao contrário do agrupamento pragmático de classes, que ocorria nos esquemas anteriores (Pontes; Lima, 2012). A teoria foi sendo aprofundada com as contribuições do *Classification Research Group* (CRG), grupo inglês fundado em 1952, a partir das discussões sobre princípios e práticas de classificação visando atender à crescente complexidade da documentação produzida na época (Pinho; Vital, 2016).

Ranganathan organizou, em 1933, sua Classificação de Dois Pontos (*Colon Classification*), também conhecida como Classificação Facetada ou Classificação Analítico-Sintética (Lima, 2002). Ela é assim chamada por empregar o sinal de dois pontos na composição de assuntos e por utilizar os princípios de análise, para a divisão de assuntos, por meio de facetas, e de síntese, na recombinação dos assuntos para classificar um documento (Pinho; Vital, 2016; Lima, 2002).

Um esquema de classificação facetado constitui-se em um conjunto mutuamente exclusivo e exaustivo de categorias combináveis, em que cada categoria é obtida a partir do isolamento de um aspecto (faceta) do objeto a ser classificado, e que, combinadas, descrevem o assunto do objeto em questão (Pontes; Lima, 2012). Ou seja, o trabalho com esquemas de classificação facetada “[...] implica simplesmente trabalhar com grupos de taxonomia de base, na qual cada documento é analisado de acordo com uma ou mais taxonomias de base ao mesmo tempo.” (Lambe, 2007, p. 37). Taxonomias de base podem ser compreendidas aqui como pequenas estruturas hierárquicas. Em um sistema de recuperação da informação, cada faceta pode ser tratada como um campo de metadados no sistema para representar e recuperar um objeto informacional.

Ranganathan apoia sua teoria em princípios que separam o domínio do conhecimento do domínio do documento, chamando de Universo do Conhecimento e Universo do Documento, respectivamente. No Universo do Conhecimento, “[...] os conceitos de um domínio são frutos de análise dos aspectos conceituais deste domínio, que se apresentam hierarquicamente arranjados, dentro de um dado

contexto, formando taxonomias” (Gomes; Campos, 2019, p. 24). Já no Universo do Documento, o assunto é apresentado como uma representação/síntese dos aspectos tratados em determinado documento, oriundos do Universo do Conhecimento (Gomes; Campos, 2019).

A partir destes dois universos, Ranganathan apresenta dois conjuntos de princípios: o Método de Faceta, para auxiliar a formação de estruturas hierárquicas no Universo do Conhecimento; e a Sintaxe Absoluta, para a representação do assunto do documento e sua respectiva notação, no âmbito do Universo do Documento. O termo faceta designa elementos que caracterizam uma ideia. Em um alto grau de abstração, é a manifestação de uma categoria (Gomes; Campos, 2019). De acordo com Pinho e Vital (2016), facetas são conjuntos resultantes da aplicação de uma determinada característica, que é definida considerando usuários, objetivos e outros elementos relevantes em determinado contexto. O desenvolvimento de uma classificação aplicando o Método de Faceta permite uma análise ampla de assuntos, propiciando simultaneidade de critérios de classificação (Pinho; Vital, 2016). O Método de Faceta se mostrou adequado à elaboração de tesouros, devido a possibilidade de diferentes tipos de relacionamentos, além da pós-coordenação dos termos. Desta forma, o Método de Faceta passa a se desvincular da construção de esquemas de classificação notacionais (Gomes; Campos, 2019). Também por este motivo, a classificação facetada tem sido adotada na organização e representação da informação em ambientes digitais, onde não é necessária notação ou arranjo físico de objetos informacionais (Pinho; Vital, 2016).

Além do Universo do Conhecimento e do Universo do Documento, a teoria de Ranganathan considera o Universo de Trabalho da Classificação, trazendo princípios e cânones para três planos de trabalho: Plano das Ideias, Plano Verbal e Plano Notacional. As ideias são as unidades de estudo presentes no Plano das Ideias. A reunião de ideias forma um assunto no Plano Verbal, representando o conteúdo de objetos informacionais por meio de termos. Os assuntos identificados no Plano Verbal são representados através de notações no Plano Notacional. Desta forma, é possível combinar diferentes ideias e características, compondo diferentes e novos assuntos, em uma indexação pós-coordenada, diferentemente de esquemas como a CDD, que apenas listavam assuntos previamente estabelecidos (Gomes; Campos, 2019).

Porém, para construir estes arranjos e combinações possíveis, foi necessário definir alguma ordem ou escala entre as ideias, e por isso, Ranganathan então buscou

algum padrão entre as ideias e foi generalizando até chegar a um número de ideias postuladas, introduzindo as Categorias Fundamentais, a saber: Personalidade, Matéria, Energia, Espaço e Tempo (PMEST) (Gomes; Campos, 2019). “As categorias fundamentais demarcam a primeira classificação de assuntos dentro de um grande universo.” (Pinho; Vital, 2016, p. 101):

Personalidade: é considerada indefinível, quando um termo não se adequa a nenhuma outra categoria, é entendida como uma manifestação desta;
 Matéria: materiais constitutivos de um objeto. Ex: Mesa/Matéria: madeira;
 Energia: processos, operações, ações, técnicas, métodos, fenômenos;
 Espaço: lugares;
 Tempo: períodos. (Pinho; Vital, 2016, p. 101).

A partir de cada Categoria Fundamental, vão se especificando os conceitos, formando facetas, apresentadas em estruturas hierárquicas. Estas estruturas são compostas por cadeias e renques. Cadeia é uma série vertical de conceitos, em que cada conceito subordinado tem uma característica a mais do que seu conceito superordenado. Renque é uma série horizontal de conceitos coordenados, formada a partir de uma única característica de divisão (Campos, 2021). Ranganathan apresenta uma série de princípios e cânones para a formação de cadeias e renques que, em geral, são o ponto de partida para elaboração de estruturas hierárquicas:

Ranganathan fornece Cânones (regras) para organizar/estruturar as entidades em um universo do conhecimento, e constituem a base teórica, enquanto os Princípios se adequam a sua aplicação em um dado ambiente. O elemento por excelência é a característica: é ela que permite a criação e organização dos conjuntos de conceitos em uma faceta. Elas são particularmente úteis na elaboração de taxonomias e tesouros, porque auxiliam na formação de classes de conceitos. (Gomes; Campos, 2019, p. 26).

A característica é um conceito fundamental no Método de Faceta. É ela que forma uma ideia e pode ser objeto de busca em um sistema de informação. Portanto, é possível pesquisar por uma característica comum a vários objetos, e não diretamente a partir de um assunto pré-estabelecido. Os cânones para características são (Gomes; Campos, 2019):

- a) Cãnone de Diferenciação: a aplicação de uma característica de divisão deve promover a diferenciação de suas entidades e produzir pelo menos duas classes;
- b) Cãnone de Relevância: a característica selecionada como critério de divisão deve ser útil à finalidade da classificação;

- c) Cãnone da Averiguabilidade ou Verificabilidade: a característica selecionada deve ser definitiva e verificável;
- d) Cãnone da Permanência: a característica selecionada deve ser mantida enquanto não houver mudança no propósito da classificação.

Uma ideia, entendida de maneira geral como um conceito, pode ter várias características. A partir da aplicação do cãnone de relevância, é possível identificar quais características serão selecionadas para gerar facetas e subfacetas. Isto permitirá que um item possa pertencer a várias classes. Ou também, dependendo do contexto, pode-se optar por classificar o objeto apenas pelo aspecto mais relevante em determinado domínio (Gomes; Campos, 2019).

Ranganathan apresenta também Cãnones para Sucessão Relevante e Consistente de características, determinando que, uma vez que são selecionadas as características para divisão, é necessário haver uma ordem entre as classes e esta ordem deve ser consistente em outras classes da mesma natureza no sistema de classificação. Na organização da informação em ambientes digitais, a ordenação é mais flexível, pois pode-se navegar e ter acesso à informação a partir de várias características de um mesmo objeto. De qualquer forma, estes cãnones são importantes para a estruturação de cadeias e renques (Gomes; Campos, 2019). Ainda em relação às características, o Cãnone de Concomitância determina que, de forma alguma duas ou mais características podem ser concomitantes em um mesmo esquema associado de características, ou seja, não devem resultar em um mesmo renque de conceitos (Campos; Gomes, 2007).

Em relação à estruturação de renques, Ranganathan define a forma como os conceitos devem ser agregados para formar classes. O Cãnone de Exaustividade determina que as classes devem esgotar todos os elementos que as compõem (Campos; Gomes, 2007). Quando aplicado em ambientes digitais, este cãnone é violado pois apenas são permitidas classes com conteúdo agregado (Gomes; Campos, 2019). O Cãnone de Exclusividade define que os elementos em um renque devem ser mutuamente exclusivos (Gomes; Campos, 2019). O Cãnone de Sequência Útil determina que a sequência dos elementos em uma classe deve ser útil para a finalidade da classificação, e o Cãnone de Sequência Consistente estabelece que a sequência dos elementos deve manter os mesmos princípios de organização em classes semelhantes (Campos; Gomes, 2007).

Em relação à estruturação de cadeias, os cânones para o Método de Faceta tratam da ordem em que cada conceito deve estar em relação aos outros conceitos (Campos; Gomes, 2007). O Cânone da Extensão Decrescente estabelece a ordem hierárquica da estrutura, partindo de classes de menor intensão para classes de maior intensão. O Cânone da Modulação define que a ordem das classes na hierarquia deve ter uma sequência que respeite todos os elos da cadeia (Campos; Gomes, 2007). Este último pode ser violado na organização da informação em ambientes digitais, caso alguma classe da cadeia não possua conteúdo atrelado (Gomes; Campos, 2019).

No tocante aos Cânones de Sequência Útil e Sequência Consistente, já relatados anteriormente, em relação à estruturação de renques, Ranganathan apresenta uma série de princípios para a ordenação dos elementos no interior de uma faceta. Estes princípios são bastante úteis para a ordenação de facetas em ambientes digitais, facilitando a estruturação da navegação na página *web*, considerando que as informações são dispostas em uma tela, de forma bidimensional (Gomes; Campos, 2019). Ranganathan elenca oito princípios (Campos; Gomes, 2007; Gomes; Campos, 2019):

- a) Princípio do Posterior-no-Tempo: ocorre principalmente em fenômenos e processos. Exemplos: escolas de pensamento, religiões, planos econômicos;
- b) Princípio do Posterior-na-Evolução: orienta a organização de conceitos relacionados a processos produtivos. Exemplo na Ciência Política: Anarquia, Primitiva, Feudal, Monarquia, Oligarquia e Democracia;
- c) Princípio da Contiguidade Espacial: refere-se a um objeto no todo, quando é necessário ordenar os elementos que o compõem, sendo apresentados em uma disposição espacial, podendo ser em uma linha direcional (vertical ou horizontal), radial, circular ou a partir de uma determinada posição;
- d) Princípio para Medida Quantitativa: ordenação por quantidade crescente ou decrescente;
- e) Princípio da Complexidade Crescente: ordenação por uma sequência de graus de complexidade;
- f) Princípio da Sequência Canônica: ordenação de acordo com tradição no domínio do conhecimento;

- g) Princípio da Garantia Literária: sequência dos itens de acordo com a quantidade decrescente de documentos publicados ou a serem publicados;
- h) Princípio da Ordem Alfabética: deve ser adotado quando nenhum outro princípio é mais relevante para a finalidade da classificação.

De acordo com Café e Bratfisch (2007), modelos baseados na lógica analítico-sintética vem ocupando maior espaço no âmbito das modernas teorias de representação do conhecimento, visto que oferecem mais possibilidades de descrição e recuperação de conteúdos multidimensionais, exigidas em Sistemas de Recuperação da Informação. Estes modelos são adequados para a representação de conteúdos semânticos complexos, através das inúmeras possibilidades de relacionamentos entre as categorias. Apesar de Ranganathan postular as cinco Categorias Fundamentais em sua teoria, Café e Bratfisch (2007, p. 239) declaram que: “Pesquisas mais recentes têm procurado demonstrar que a análise de um determinado assunto pode ir além dessas categorias, o que não invalida seu valor teórico e metodológico para o desenvolvimento da pesquisa sobre classificação.”

Originalmente a Teoria da Classificação Facetada destinava-se a fornecer notações para ordenação de livros em estantes de bibliotecas. Porém, sua teoria ultrapassou esta finalidade inicial, indo além da representação de assuntos em documentos, mostrando sua relevância para a descrição de objetos digitais diversos (Gomes; Campos, 2019). Atualmente a classificação facetada é empregada na organização da informação em ambientes digitais, oferecendo uma base para a definição de metadados para indexação e recuperação da informação, através de busca, navegação e aplicação de filtros, a partir de diferentes perspectivas do mesmo objeto. A classificação facetada tem espaço na organização da informação na *web*, que por ser um ambiente virtual, permite mais facilmente apresentar a mesma informação organizada de várias formas (Carlan; Medeiros, 2011).

De acordo com Pontes e Lima (2012, p. 28):

O formato lógico e previsível da classificação facetada se torna bastante intuitivo e útil ao usuário, podendo ser usada na construção da interface de navegação e exploração do acervo, como alternativa à estrutura hierárquica de diretórios típica das classificações tradicionais. As classificações facetadas podem fornecer pontos de acesso multidimensionais flexíveis e altamente estruturados aos documentos.

As facetas em uma classificação analítico-sintética podem combinar aspectos semânticos e descritivos de documentos ou objetos. Em relação a este aspecto, Pontes e Lima (2012) ressaltam que atualmente tem se mostrado uma visão diferenciada sobre o conceito de facetas na literatura, podendo incorporar aspectos não semânticos de um documento. A norma ANSI/NISO Z39.19:2005 (National Information Standards Organization, 2010, p. 14, tradução nossa) afirma que: “A análise facetada é, às vezes, usada para indicar atributos de objetos de informação”⁹. Ou seja, uma entidade ou objeto, e não um assunto, é classificado de acordo com suas propriedades (elementos descritivos). Neste caso, cada propriedade representa uma faceta que agrupa um conjunto de instâncias (entidades). Estas facetas podem ser facilmente representadas por meio de algum padrão de metadados em um sistema de informação (Pontes; Lima, 2012).

Esta pesquisa busca desenvolver um modelo de representação de produtos do varejo comercializados em plataformas digitais. Portanto, esta última abordagem sobre facetas parece ser mais adequada ao contexto desta pesquisa, visto que o produto é um objeto e não um assunto, ou seja, não carrega em si um conteúdo informacional. Sendo assim, revela-se ser mais compatível sua classificação a partir de seus atributos, por meio de um padrão de metadados para sua representação formal.

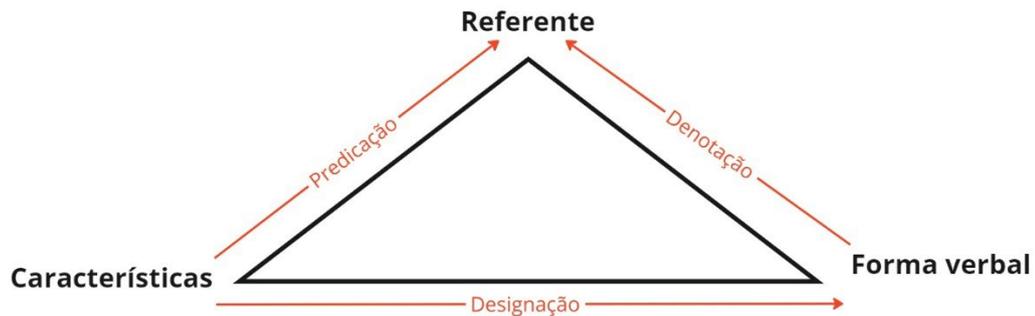
2.1.2.2 Teoria do Conceito

Outra base teórica para o desenvolvimento de SOCs é a Teoria do Conceito, “[...] desenvolvida por Ingetraut Dahlberg na década de 1970, que viabiliza uma fundamentação sólida para a determinação e o entendimento dos conceitos, assim denominados, com a finalidade de representação e recuperação da informação.” (Carlan; Medeiros, 2011, p. 60).

De acordo com Dahlberg (1978a), a identificação de um conceito depende da interação entre três elementos, sendo eles: o Referente, as Características e a Forma verbal, que podem ser observados na figura abaixo:

Figura 1 – Triângulo Conceitual de Dahlberg

⁹ No original: “Facet analysis is sometimes used to indicate the attributes of content objects” (National Information Standards Organization, 2010, p. 14).



Fonte: Adaptado de Dahlberg (1978a).

No triângulo conceitual de Dahlberg, as características são atribuídas a um item de referência e são sintetizadas em uma forma verbal. Se um dos elementos do triângulo não estiver presente, não pode haver o conceito. O conceito, portanto, é uma unidade do conhecimento que compreende afirmações verdadeiras sobre um determinado item de referência, representado em uma forma verbal (Gomes; Campos, 2019). Ou seja, as características são formadas por um conjunto de afirmações consideradas como verdadeiras por membros de uma área do conhecimento, sobre um item de referência (predicação). A soma dos enunciados verdadeiros sobre um referente fornece o conceito deste (Dahlberg, 1978b). As características são sintetizadas em uma forma verbal por meio de termos e definições (designação). Por sua vez, a forma verbal representa o conceito por meio de um signo linguístico (denotação) (Gomes; Campos, 2019).

Este triângulo e a interação entre estes três elementos evidenciam a abordagem onomasiológica da Teoria do Conceito, uma vez que atua no plano extralinguístico para chegar no plano linguístico (termo). Os termos são, portanto, apenas a aparência externa de algo que é representado como um conceito (Gomes; Campos, 2019). A abordagem onomasiológica parte do referente e suas características para chegar no termo, portanto é considerada uma abordagem conceitual, “[...] que parte de uma perspectiva sistêmica na qual o conjunto dos termos de um domínio especializado reflete a organização do conhecimento nesse domínio, ou seja, os termos denotam os conceitos.” (Gomes; Campos, 2019, p. 38).

O conceito do ponto de vista do trabalho terminológico é considerado como uma representação mental de objetos em um determinado contexto ou área especializada. Objeto aqui é compreendido como um objeto individual existente no mundo e que, através da observação e por um processo de conceitualização, é categorizado em classes, que correspondem a unidades do conhecimento, ou seja,

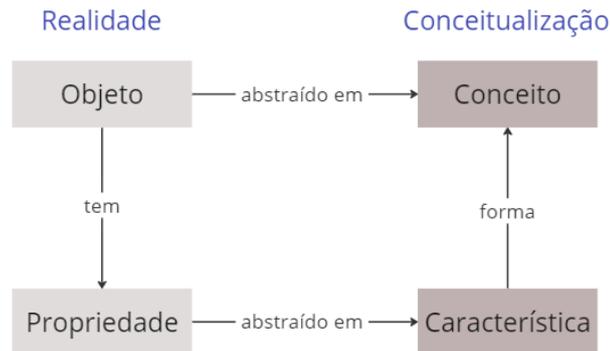
conceitos. Os conceitos são divididos como conceitos gerais ou individuais. Os conceitos gerais correspondem a um conjunto de dois ou mais objetos, formando um grupo com propriedades em comum (International Organization for Standardization, 2022). Em um SOC, são compreendidos como classes e são designados por termos. Os conceitos individuais retratam um único objeto, distinto dos demais, em um determinado espaço e tempo (Dahlberg, 1978b). Em SOCs, são considerados como instâncias de classes e são representados por uma denominação, como por exemplo “Universidade Federal do Rio Grande do Sul”. De acordo com Dahlberg (1978b, p. 102): “É fácil também verificar que os elementos contidos nos conceitos gerais se encontram também nos conceitos individuais, sendo, portanto, possível reduzir os conceitos individuais aos gerais e ordená-los de acordo com os conceitos gerais.”

Para o desenvolvimento de sistemas, os conceitos podem ser categorizados formalmente conforme lista abaixo. Esta categorização se aproxima às facetas e categorias fundamentais, trazidas por Ranganathan, e podem ser combinadas de inúmeras maneiras.

- A Objetos. Ex.: plantas, produtos, papel, etc.
- B Fenômenos. Ex.: crescimento, chuva, tráfego, etc.
- C Processos. Ex.: imprimir, sintetizar, etc.
- D Propriedades. Ex.: cego/cegueira, suave/suavidade, etc.
- E Relações. Ex.: causalidade, necessidade.
- F Dimensão. Ex.: espaço, tempo, posição, etc. (Dahlberg, 1978b, p. 105).

A formação do conceito desempenha papel fundamental na organização do conhecimento. No trabalho terminológico, para a formação de conceitos, é necessário identificar as propriedades dos objetos. Os objetos que compartilham as mesmas propriedades são agrupados em unidades. Se este grupo de objetos é considerado significativo no domínio de conhecimento, então esta é uma unidade significativa, e as propriedades em comum que os objetos desta unidade significativa compartilham são abstraídas como características, que são combinadas em um conjunto na formação do conceito. Por este motivo, os objetos no mundo real são identificados por propriedades. Os objetos são abstraídos como conceitos e as propriedades são abstraídas como características que formam o conceito (International Organization for Standardization, 2022). Isto pode ser observado na figura abaixo:

Figura 2 – Processo de conceitualização



Fonte: Adaptado de International Organization for Standardization (2022).

No trabalho terminológico, são identificados objetos no contexto do domínio de conhecimento em questão, a partir da análise de textos discursivos. Com isso, são identificadas as propriedades destes objetos para então determinar as propriedades que podem ser abstraídas como características, ou seja, que não são únicas de um objeto individual. Para facilitar esta análise, as propriedades podem ser categorizadas como parte, função, composição, cor, formato, operação e localização. Porém é importante ter flexibilidade quanto a estas categorias, considerando demais categorias apropriadas a um determinado domínio do conhecimento, que podem ser encontradas em obras de referência (International Organization for Standardization, 2022). Dahlberg (1978b) indica uma ordem diferente de categorização de características, com base nas categorias aristotélicas: matéria (substância), qualidade, quantidade, relação, processo (atividade), modo de ser, passividade, posição, localização e tempo.

A soma total de características, configura a intensão de um conceito. Na intensão, também estão compreendidos os conceitos superordenados e as características delimitadoras, que diferenciam um conceito. Já a extensão do conceito é compreendida como a soma total dos seus conceitos mais específicos. “Pode também ser entendida como a soma dos conceitos para os quais a intensão é verdadeira, ou seja, a classe dos conceitos de tais objetos dos quais se pode afirmar que possuem aquelas características em comum que se encontram na intensão do mesmo conceito.” (Dahlberg, 1978b, p. 105). A extensão é dividida em duas espécies (Dahlberg, 1978b):

- a) extensão de um conceito genérico em relação aos seus conceitos específicos. Exemplo:

casa

casa de pedra

casa de madeira

b) extensão de um conceito genérico em relação a seus objetos. Exemplo:

casa

casa do Presidente da República

casa do vizinho

Pode-se observar que a intensão determina a extensão do conceito, pois todos os objetos e conceitos específicos relacionados a um conceito geral devem apresentar as características que constituem a intensão deste conceito geral.

Após a identificação das características que compõem a intensão do conceito, no trabalho terminológico, cada característica é analisada em relação aos demais conceitos no sistema conceitual. São identificadas características em comum entre conceitos, chamadas de características compartilhadas, e as diferenças que separam um conceito são identificadas como características delimitadoras. Este conjunto de características tipifica um determinado conceito e o situa em um sistema conceitual, relacionando-o a conceitos com características semelhantes ou diferentes (International Organization for Standardization, 2022).

As características podem ser divididas entre simples e complexas. São consideradas simples as que se referem a uma única propriedade. São complexas as características que dizem respeito a mais de uma propriedade. As características podem ser ordenadas como essenciais ou acidentais, para a constituição de um conceito (Dahlberg, 1978b).

a) características essenciais (necessárias):

- características essenciais constitutivas;
- características essenciais consecutivas;

b) características acidentais (adicionais ou possíveis):

- características acidentais gerais;
- características acidentais individualizantes.

As características essenciais consecutivas dependem das características essenciais constitutivas (Dahlberg, 1978b). De acordo com Gomes e Campos (2019), as características essenciais são elementos necessários e suficientes, em um determinado contexto, para determinar o tipo do conceito, ou seja, sua identidade. Portanto, as características essenciais têm a função de estruturar sistemas de classificação (Gomes; Campos, 2019). As características acidentais dependem de fatores externos e de condições acidentais (Dahlberg, 1978b). Elas não são necessárias ou constantes, são opcionais para um determinado tipo de conceito

(Gomes; Campos, 2019). Em relação a produtos, objeto de estudo deste trabalho, Dahlberg (1978b) afirma que a intensão do conceito é determinante, sendo formada por características essenciais determinadas pelas finalidades e aplicação do produto. Neste contexto, as características acidentais dependem da respectiva eficiência ou outros valores práticos.

A norma ISO 704 (International Organization for Standardization, 2022) estabelece que as características são usadas para a análise conceitual, na modelagem de sistemas conceituais, na formulação de definições e pode auxiliar na formação de designações (termos). Esta afirmação vai ao encontro de Dahlberg (1978b), que declara que as características servem para a ordenação classificatória dos conceitos, definição e formação dos nomes dos conceitos. No processo de identificação de características de conceitos, caso dois os mais conceitos apresentem pelo menos uma característica em comum, oposta ou semelhante, é possível admitir que entre eles são estabelecidas relações (Dahlberg, 1978b; Gomes; Campos, 2019). “Um conjunto de conceitos estruturados de acordo com as relações entre eles é chamado de sistema conceitual.” (International Organization for Standardization, 2022, p. 8, tradução nossa).¹⁰

As relações conceituais podem ser divididas em três grupos: relação categorial formal, relação lógico formal e relação semântica. As relações categoriais formais se dão quando há uma hierarquia de características, chegando a características gerais que se tornam categorias (Carlan; Medeiros, 2011). Na modelagem de um domínio, as relações categoriais são fundamentais para uma primeira estruturação do conceito no interior de um domínio (Gomes; Campos, 2019). Gomes e Campos (2019) afirmam que Dahlberg parte de quatro categorias formais, sendo estas: entidades, qualidades, atividades e dimensões. Elas se aproximam das categorias fundamentais de Ranganathan, conhecidas como PMEST (personalidade, matéria, energia, espaço e tempo).

As relações lógicas formais medem a quantidade e a similaridade das características de um conceito, sendo subdivididas em cinco tipos (Gomes; Campos, 2019; Carlan; Medeiros, 2011; Dodebei, 2002; Dahlberg, 1978b): identidade, implicação, interseção, disjunção e negação. A relação de identidade ocorre quando dois conceitos apresentam características semelhantes. Do ponto de vista

¹⁰ No original: “A *concept system* shall be structured according to the relations between the *concepts*.” (International Organization for Standardization, 2022, p. 8).

terminológico, se estabelece uma relação de sinonímia, em que um conceito pode ser representado por mais de um símbolo distinto. Na aplicação em tesouros, estabelecem-se as relações de equivalência para controle de dispersão léxica (sinônimos e quase sinônimos), dispersão simbólica (grafias diferentes, abreviaturas, tradução) e dispersão sintática (coordenação, gênero e número).

A relação de implicação, chamada também de inclusão, ocorre quando, entre dois conceitos, um deles apresenta uma característica adicional, ou seja, um conceito está contido em um outro conceito. Já a relação de interseção ocorre quando características entre dois conceitos se entrecruzam, ou seja, os conceitos coincidem em alguma característica. A relação de disjunção ocorre quando dois conceitos são mutuamente exclusivos, não apresentando nenhuma característica em comum. Por último, a relação de negação ocorre quando um conceito inclui uma característica cuja negação está contida em outro conceito.

As relações semânticas podem ser divididas em paradigmáticas e sintagmáticas (Carlan; Medeiros, 2011; Gomes; Campos, 2019; International Organization for Standardization, 2022; Dodebei, 2002; Dahlberg, 1978b; National Information Standards Organization, 2010). As relações paradigmáticas ocorrem entre conceitos na mesma categoria e podem ser divididas em: hierárquicas, partitivas e de oposição. A relação hierárquica, conhecida também como relação de abstração, genérica ou gênero/espécie, ocorre quando dois conceitos apresentam as mesmas características, ou seja, a mesma intensão, e um deles possui uma característica delimitadora adicional, criando níveis de conceitos superordenados e subordinados, em que estes últimos herdam as características dos conceitos superordenados. A característica adicional é selecionada como critério de subdivisão do conceito. Quando o conceito é subdividido por mais de um critério de subdivisão, o sistema conceitual resultante é chamado de multidimensional ou facetado. Do ponto de vista da lógica, aplica-se na relação hierárquica a relação de implicação.

A relação partitiva, chamada também de relação de existência ou todo-parte, consiste na decomposição de um todo (item de referência) em suas partes, em que o todo inclui as características de suas partes. Apesar de haver também uma hierarquia entre o todo e suas partes, os conceitos subordinados não herdam as características do conceito superordenado. Na relação partitiva, uma ou mais partes podem ser obrigatórias ou opcionais, dependendo do item de referência em questão. Caso uma parte pertença a mais de um todo, indica-se não estabelecer uma relação hierárquica

entre a parte e o todo, mas sim, uma relação associativa, para evitar classificação cruzada. Por exemplo, uma roda pode ser parte de uma motocicleta ou de um caminhão. De acordo com Dahlberg (1978b), a relação entre um produto e os elementos que o constituem configura-se como uma relação partitiva.

A relação de oposição, conhecida também como relação de complementaridade, aplica a relação lógica de negação. Em um tesouro, podem aparecer nas relações hierárquicas ou associativas. Podem ser divididas da seguinte forma:

- a) oposição contraditória: numérico/não numérico, presença/ausência;
- b) oposição contrária: branco/preto, amizade/inimizade;
- c) oposição positivo – neutro – negativo (gradação): favorável-neutro-desfavorável, alto-igual-baixo.

Além das relações paradigmáticas, as relações sintagmáticas também são caracterizadas como relações semânticas. Elas não são hierárquicas, ocorrendo entre conceitos em diferentes categorias, devido a alguma conexão temática em virtude da experiência. Do ponto de vista da lógica, aplica-se aqui a relação de interseção. Em um tesouro, são representadas por meio de relações associativas. Em sua maioria, são relações funcionais, que estabelecem uma conexão entre um objeto e as atividades e/ou funções relacionadas no mundo real, por meio de valências semânticas dos verbos. Por exemplo: para o verbo “produzir”, pode-se questionar quem produz (produtor), o que é produzido (produto) ou como é produzido (produção).

Gabriel Junior e Laipelt (2019) apresentam proposta de metacategorias de relações associativas para enriquecimento semântico de tesouros, com base na norma ISO 25964 (International Organization for Standardization, 2011), conforme segue: causa/efeito, afinidade, características do produto, agente/processo, ação/produto da ação, ação/paciente ou objetivo, ação/propriedade, campo de estudo/objetos ou fenômenos estudados, ação ou coisa/contra agente, conceito ou coisa/propriedades, coisa/suas partes, matéria prima/produto, antonímia e oposição conceitual, coordenação. A norma ISO 25964 (International Organization for Standardization, 2011) ressalta a importância da explicitação das relações associativas, utilizando algum tipo de convenção.

De acordo com Lima (2020, p. 21), “as relações semânticas são fundamentais para a compreensão de representações da realidade, pois facilitam a assimilação pelo usuário do propósito da associação entre os conceitos no contexto que lhe é

apresentado.” Este contexto é influenciado pela linguagem e cultura, pelo conhecimento de mundo do usuário e pelo domínio do conhecimento (Lima, 2020).

O conceito é representado de forma verbal em um sistema conceitual por meio de um termo e sua definição. O termo é uma forma sucinta de referenciar um conceito, enquanto a definição deve possibilitar escolher a extensão do conceito e distingui-lo de outros dentro do mesmo sistema conceitual (International Organization for Standardization, 2022). Dahlberg (1978b, p. 106) afirma que existe consenso que as “[...] definições são pressupostos indispensáveis na argumentação e nas comunicações verbais e que constituem elementos necessários na construção de sistemas científicos.” Definição é a “[...] delimitação ou fixação do conteúdo de um conceito (conteúdo do conceito = intensão, ou conjunto de características ou atributos).” (Dahlberg, 1978b, p. 106). A definição é apresentada como uma equação de sentido, em que, do lado esquerdo, encontra-se aquilo que deve ser definido (*definiendum*) e do lado direito, aquilo pelo qual algo é definido (*definiens*) (Dahlberg, 1978b).

As definições podem ser categorizadas como nominais ou reais. As definições nominais têm por finalidade a fixação de sentido de uma palavra, da forma que ela ocorre na linguagem (Dahlberg, 1978b). Neste sentido, o conteúdo da definição é uma equivalência textual do termo (Gomes; Campos, 2019). Já as definições reais, conhecidas também por definições conceituais, delimitam a intensão de um determinado conceito, distinguindo-o dos demais, relacionando-se com o conhecimento contido em determinado conceito, em um domínio do conhecimento. Na definição real, são mencionadas as características essenciais de um referente e, por vezes, as características acidentais. As definições reais podem ser divididas em simples ou complexas. Em ambas, encontra-se na definição um conceito mais amplo do que está contido no termo, sendo seguido de uma característica delimitadora. Na definição real complexa, o termo e a diferença específica são acompanhados de suplementos (Dahlberg, 1978b).

A norma ISO 704 (International Organization for Standardization, 2022) indica a definição conceitual, denominada pela norma como definição intensional, como o método de definição de conceito mais explícito e preciso. Uma definição conceitual define um conceito como uma unidade com uma intensão inequívoca refletida por uma extensão única. Ela deva fornecer uma quantidade mínima de informação para que

seja possível reconhecer e diferenciar um conceito de outros relacionados, especialmente entre conceitos coordenados.

As definições conceituais devem incluir o conceito imediatamente superordenado, sendo seguido pelas características delimitadoras. Com isso, as características que formam a intensão do conceito superordenado são assumidas implicitamente na definição do conceito subordinado, devido ao princípio da herança. Por isso, é importante que os conceitos superordenados sejam definidos antes de seus conceitos subordinados. As definições devem ser tão concisas quanto possível e tão complexas quanto necessário, e devem ser formuladas de maneira consistente, levando em consideração a linguagem do público-alvo e seu nível de conhecimento (International Organization for Standardization, 2022).

Gomes e Campos (2019, p.40) afirmam que:

[...] o ato de definir é uma construção de aspectos selecionados do referente realizado pelo “olhar” de um grupo que aquiesce sobre o seu entendimento em um dado espaço de conhecimento. Desta forma, esses elementos que formam o conceito, as características são propriedades dos referentes (objetos concretos ou abstratos) que, no nível do conceito, passam a ser também características do conceito.

Esta afirmação vai ao encontro da indicação da norma ISO 704 (International Organization for Standardization, 2022), que declara que é importante considerar as necessidades do público-alvo ao fornecer informações sobre conceitos. Por isso, deve-se consultar especialistas no domínio do conhecimento em questão, tanto especialistas familiarizados com os padrões de conceituação na área, quanto especialistas menos familiarizados, que podem ser de outra área do conhecimento, além de não especialistas.

Em suma, a Teoria do Conceito oferece um método científico para análise de um item de referência em um determinado domínio, por meio de suas propriedades, a partir de uma abordagem onomasiológica. Com isso, submete as propriedades do objeto a um processo de abstração, em que se identificam as características dos conceitos. Este processo é de fundamental importância, em qualquer esfera de organização do conhecimento, para a compreensão de um conceito em um determinado domínio, possibilitando sua ordenação no interior deste domínio, ou seja, construindo um sistema conceitual que represente o domínio do conhecimento (Gomes; Campos, 2019).

2.1.2.3 Terminologia

Outra base teórica em que se pode amparar o desenvolvimento de SOC é a Terminologia. A Terminologia como disciplina é definida como “[...] estudo e tratamento de termos técnicos e/ou científicos.” (Krieger, 2008, p. 1). A Terminologia é uma das disciplinas que compõem as ciências do léxico, ao lado da Lexicologia e da Lexicografia. Seu objeto de estudo são as unidades lexicais, porém em âmbito especializado, ou seja, em domínios especializados do conhecimento (Krieger, 2008).

De acordo com Carlan e Medeiros (2011, p. 65), “nos contextos técnico-científicos, a terminologia assume papel fundamental por trabalhar com as línguas de especialidades, responsáveis por garantir a comunicação rápida e precisa entre os interlocutores.” A transferência de informação, especialmente na área científica, requer um sistema de significação claro e explícito, buscando evitar ambiguidades (Carlan; Medeiros, 2011).

Na Terminologia, o termo representa um conceito que está confinado a um domínio do conhecimento, cujo conteúdo é expresso em uma linguagem especializada. Portanto, a aderência a um domínio do conhecimento é o que separa a linguagem comum da linguagem especializada (Warburton, 2021). Neste sentido, é possível afirmar que existe linguagem especializada em organizações comerciais e corporativas, contexto desta pesquisa, visto que estes ambientes possuem um número limitado de usuários (definidos por profissão ou especialização), estão inseridos em uma situação comunicativa formal ou profissional e apresentam uma função informativa, caracterizando, assim, uma linguagem especializada (Cabré, 1999b).

O estudo das teorias da Terminologia pode explicar como termos, conceitos e relações semânticas devem ser selecionados e relacionados na constituição de um SOC (Carlan; Medeiros, 2011). Portanto, “[...] estudos atuais na organização e representação do conhecimento recomendam a escolha de uma das teorias da Terminologia para a construção de SOC, de acordo com o campo do conhecimento a ser trabalhado.” (Carlan; Medeiros, 2011, p. 65).

As principais teorias da Terminologia são a Teoria Geral da Terminologia (TGT), Teoria da Socioterminologia e a Teoria Comunicativa da Terminologia (TCT) (Carlan; Medeiros, 2011).

A Teoria Geral da Terminologia (TGT) foi elaborada por Eugen Wüster em sua tese de doutorado, em 1931, e foi pioneira e responsável pela base do trabalho terminológico (Carlan; Medeiros, 2011; Laipelt; Monteiro-Krebs, 2021). A TGT privilegiou a dimensão conceitual do termo, em oposição ao aspecto linguístico, caracterizando o termo como uma unidade cognitiva, dissociando a forma de expressão (termo) do conteúdo (conceito) (Krieger, 2008). Ou seja, o termo é analisado a partir do conceito que ele representa, que segundo Cabré (1999a) demonstra o aspecto reducionista da TGT, remetendo à concepção de universalidade dos conceitos nos domínios especializados, sendo diametralmente oposto ao conceito de variação (Laipelt; Monteiro-Krebs, 2021). Em suma, a TGT possui uma perspectiva prescritiva e normativa que visa rotular e padronizar o uso de termos e conceitos, sem considerar as inúmeras variações possíveis em um contexto comunicacional (Carlan; Medeiros, 2011).

Apesar de ser uma teoria defasada frente a teorias mais modernas da Terminologia, que estudam os termos especializados da perspectiva comunicativa e linguística, o estudo de Wüster sobre a natureza dos conceitos é importante para a área de OC. Por exemplo, a concepção de intensão e extensão conceitual são muito utilizadas para a elaboração de SOCs e se relacionam com a Teoria do Conceito, de Dahlberg (1978b).

A partir da década de 1990, houve um crescimento no interesse dos linguistas sobre a linguagem em ambientes profissionais, especialmente devido ao avanço da tecnologia e da popularização das ciências (Serra, 2019), o que levou ao estudo sobre variação terminológica. A primeira teoria a trazer o aspecto da variação foi a Socioterminologia, formalizada em 1993, por François Gaudin (Laipelt; Monteiro-Krebs, 2021). A Socioterminologia apresenta uma atitude descritiva, se contrapondo à abordagem prescritiva da TGT, pois considera metáforas e o contexto histórico de uso das palavras, questionando a monosssemia dos termos (Lara, 2006).

A Teoria Comunicativa da Terminologia (TCT) foi desenvolvida por Maria Teresa Cabré, no final da década de 1990, e compreende as unidades terminológicas como parte da linguagem natural e da gramática das línguas, valorizando os aspectos comunicativos das linguagens de especialidade, a partir da dimensão textual e discursiva dos termos, e admitindo a variação terminológica (Laipelt; Monteiro-Krebs, 2021). Ou seja, os termos não pertencem a um domínio específico do conhecimento, eles fazem parte do léxico comum e assumem a função de termo especializado de

acordo com o contexto textual e discursivo em que ocorrem. Na TCT, o termo é considerado uma unidade complexa, poliédrica, abrangendo três ângulos básicos: o linguístico, o cognitivo e o comunicacional (Cabré, 1999a).

Como fundamento teórico para esta pesquisa, será adotada a TCT, devido à sua abordagem descritiva e que contempla a variação terminológica, visto que para o desenvolvimento de um modelo de representação, é importante considerar os conceitos e os termos como ocorrem no domínio específico do conhecimento, incluindo suas variações.

De acordo com Cabré (1999a), na situação específica do discurso especializado, os termos se apresentam como itens lexicais que adquirem “conjuntos de características específicas”, e por meio desses conjuntos eles obtêm o status de termo. De acordo com a sua utilização em um contexto específico, são ativados módulos de características apropriados para este contexto, assumindo o caráter de termo neste contexto. “Os cenários comunicativos fornecem, portanto, as chaves de acesso à identificação dos mecanismos de ativação do valor especializado que tanto unidades lexicais, quanto estruturas fraseológicas adquirem nas comunicações especializadas.” (Krieger, 2008, p. 4).

Laipelt e Monteiro-Krebs (2021) declaram que os termos têm duas funções: a representação e a transferência do conhecimento: “[...] um termo representa conceitos de uma área de especialização, transmite conhecimento especializado e possui caráter linguístico em função de sua inserção no discurso.” (Laipelt; Monteiro-Krebs, 2021, p. 34-35). Os termos, representando os conceitos de uma área especializada, mantêm entre si relações que constituem uma estrutura conceitual, que representa um domínio do conhecimento (Lara, 2006). Por isso a importância do uso de bases teóricas terminológicas para o desenvolvimento de SOCs.

De acordo com a norma ISO 704 (International Organization for Standardization, 2022), termos “[...] representam conceitos gerais que correspondem a vários tipos de objetos, por exemplo, coisas, atividades, situações ou relações.” (International Organization for Standardization, 2022, p. 49, tradução nossa).¹¹ Esta norma destaca que quando é necessária uma comunicação precisa e exata em uma linguagem especial, é importante alcançar a monossêmia, pelo menos dentro de um

¹¹ No original: “[...] represent *general concepts* that correspond to various types of *objects*, e.g. things, activities, situations or relations.” (International Organization for Standardization, 2022, p. 49).

domínio do conhecimento, buscando reduzir a ambiguidade (International Organization for Standardization, 2022).

Esta concepção parece se aproximar da TGT, pelo caráter prescritivo, o que é compreendido pela necessidade de controle terminológico no desenvolvimento de linguagens especiais. Assim como ocorre na elaboração de vocabulários controlados, em que se utiliza as duas concepções da terminologia: diversificada, contemplando as variações terminológicas, mais próxima da abordagem da TCT; e normativa, necessária para controlar a variação, buscando evitar ambiguidades na recuperação da informação, se aproximando da abordagem da TGT (Laipelt; Monteiro-Krebs, 2021). Portanto, na elaboração de SOCs, do ponto de vista terminológico, é importante aceitar e contemplar a variação, mas de forma controlada, para que não comprometa a consistência do instrumento e para que não gere ambiguidade na recuperação da informação.

As variações terminológicas tratam-se de diferentes manifestações lexicais de um conceito em um campo especializado (Tercedor Sánchez, 2013). Elas são fenômenos comunicativos que proporcionam a diversidade linguística no contexto especializado. Portanto, fazem parte do processo comunicativo, adaptando o discurso de especialistas de acordo com o nível de abstração de seus interlocutores (Laipelt; Monteiro-Krebs, 2021).

Cabré (1999b) considera como parâmetros da variação, o grau de abstração e os propósitos comunicativos: “O primeiro parâmetro depende da temática em si, dos destinatários e das funções comunicativas do emissor. Já o segundo parâmetro determina os tipos de texto, segundo as funções que lhe cabem.” (Laipelt; Monteiro-Krebs, 2021, p. 58). Cabré (1999b) também destaca dois aspectos complementares que podem influenciar as variações em linguagens de especialidade: o estilo pessoal do especialista e os dialetos geográficos, históricos e sociais.

Os estudos da TCT demonstram que as variações podem ocorrer em duas dimensões: a dimensão conceitual e a dimensão denominativa. A variação conceitual “[...] refere-se à diferenciação na conceitualização do termo, dependendo também de fatores externos à definição.” (Laipelt; Monteiro-Krebs, 2021, p. 58). Este fenômeno normalmente ocorre na adaptação do discurso especializado para diferentes níveis de abstração.

Já a variação denominativa é identificada pelas diferentes formas pelas quais pode-se representar um conceito, podendo receber diferentes designações para o

mesmo conceito, em uma área do conhecimento (Freixa, 2002). Em vocabulários controlados, como tesouros, quando há variação denominativa, uma das designações do conceito é escolhida como termo preferido e as variantes são incluídas no sistema como termos equivalentes, por meio de remissivas, possibilitando a recuperação da informação pelas diferentes designações apresentadas para o mesmo conceito (Laipelt; Monteiro-Krebs, 2021).

A escolha por utilizar uma designação ou outra para representar um conceito resulta de uma motivação comunicativa, “[...] como uma intenção particular do autor, um texto para uma área geográfica específica ou em um registro específico, ou ser causada por razões de natureza cognitiva (consciente ou não) que impliquem que uma característica ou dimensão específica do conceito seja focalizada.” (Tercedor Sánchez, 2013, p. 81, tradução nossa).¹²

Freixa (2002) elaborou proposta que visa organizar as causas da variação denominativa, dividindo em seis grupos (Serra, 2019, p. 194-195):

- a) causas prévias: redundância linguística, arbitrariedade do signo linguístico, possibilidade de variação nas línguas naturais;
- b) causas dialetais: variação diatópica, cronológica e social;
- c) causas funcionais: adequações ao nível da língua e ao nível de especialização;
- d) causas discursivas: evitar repetições, economia linguística, criatividade, ênfase e expressividade;
- e) causas interlinguísticas: convivência entre termos de línguas diferentes e oriundos de grupos de pesquisa de países e continentes diferentes;
- f) causas cognitivas: imprecisão conceitual, distância ideológica e diferença de conceptualização.

De acordo com Freixa (2013, p. 43, tradução nossa): “As causas prévias e as causas cognitivas são de fato a origem da variabilidade e explicam as alternâncias denominativas causadas pelas causas dialetais, funcionais, estilísticas e interlinguísticas [...]”¹³

A variação dialetal refere-se à diversificação denominativa de um conceito entre especialistas de países diferentes (Laipelt; Monteiro-Krebs, 2021), ou até mesmo, as variações devido a regionalismos dentro de um mesmo país, por exemplo, os termos “bergamota” e “mexerica”. A variação funcional ocorre internamente, quando o mesmo

¹² No original: “[...] como una intención particular en el autor, un texto para una zona geográfica determinada o en registro concreto, o estar provocada por razones de orden cognitivo (conscientes o no) que implican que se focalice una característica o dimensión concreta del concepto.” (Tercedor Sánchez, 2013, p. 81).

¹³ No original: “Las causas previas y las causas cognitivas de hecho son el origen de la variabilidad y explican las alternancias denominativas provocadas por las causas dialectales, funcionales, estilísticas y las interlingüísticas [...]” (Freixa, 2013, p. 43).

falante usa diferentes designações dependendo do contexto, por exemplo, o nome científico “cefaleia” e o nome popular “dor de cabeça”. As variações discursivas referem-se a questões estilísticas no discurso, especialmente para evitar repetições do mesmo termo no texto. Já as variações interlinguísticas, referem-se aos empréstimos terminológicos entre diferentes línguas.

A norma ISO 704 (International Organization for Standardization, 2022) apresenta as relações entre o termo e o conceito, que podem gerar variações terminológicas, conforme abaixo:

- a) monossêmia: ocorre quando um termo representa apenas um conceito;
- b) sinonímia: capacidade de mais de um termo representar o mesmo conceito. Esse fenômeno pode ocorrer na utilização de termos com campos semânticos próximos, ou quando duas palavras de origens diferentes entram no léxico de uma língua como empréstimos, e possuem o mesmo significado (Schiessl; Shintaku, 2012). Se dois termos são designados a conceitos que apresentam intensão quase idêntica, eles são chamados de quase-sinônimos e podem ser intercambiáveis apenas em algumas situações;
- c) antônimos: termo designa conceito contrário ou contraditório a outro conceito;
- d) polissemia: ocorre quando termos têm forma idêntica, tanto fonética quanto escrita, mas designam conceitos diferentes, mas que guardam algum sentido em comum entre eles. Em muitos casos, o mesmo termo possui acepções diversas em áreas do conhecimento distintas (Schiessl; Shintaku, 2012). Exemplo: “ponte” (tratamento dentário) e “ponte” (construção);
- e) Homonímia: ocorre quando termos têm forma idêntica, porém se referem a conceitos diferentes que não guardam nenhuma característica em comum. Exemplo: “manga” (camisa) e “manga” (fruta) (Schiessl; Shintaku, 2012). A homonímia é dividida em três situações (International Organization for Standardization, 2022):
 - homófonos: termos foneticamente idênticos, mas ortograficamente diferentes, representando conceitos não relacionados;
 - homógrafos: termos ortograficamente idênticos, mas foneticamente diferentes, representando conceitos não relacionados;

- homônimos: são termos escritos e pronunciados da mesma maneira, representando conceitos não relacionados.

A norma ISO 704 (International Organization for Standardization, 2022) apresenta alguns princípios que podem ser utilizados ao criar ou definir novos termos. O primeiro deles é o princípio da transparência, que afirma que o termo é transparente quando é possível inferir seu significado sem necessidade de haver uma definição ou outra informação suplementar. Para o termo ser transparente, normalmente uma característica do seu conceito é expressa no próprio termo. O segundo princípio é o da consistência e diz que o termo deve ser integrado e ser consistente com o sistema conceitual existente. O terceiro princípio trata da adequação, em que os termos devem seguir os padrões de significado que são familiares e estabelecidos no contexto específico. Também os termos devem ser tão neutros quanto possíveis, evitando distração ou conotações negativas (International Organization for Standardization, 2022).

O princípio da concisão trata da economia linguística, evitando que o termo se torne muito longo. Este princípio pode conflitar com os princípios de transparência e adequação, porém, nestes casos, a praticidade deve guiar a decisão pelo termo a ser adotado. O princípio da derivabilidade e composibilidade se refere a formações produtivas de termos que permitem derivados e compostos. O princípio da correção linguística determina que os termos devem estar de acordo com as normas morfológicas, morfossintáticas e fonológicas da língua em questão. Por último, o princípio da preferência pela língua nativa diz que as expressões da língua nativa devem ter preferência sobre empréstimos linguísticos, apesar de admitir o empréstimo de outras línguas como uma forma aceita para formação de termos (International Organization for Standardization, 2022).

A partir destes pontos, observa-se como a Terminologia é fundamental para instrumentalizar teórica e metodologicamente o processo de controle terminológico, que é um elemento comum e necessário a qualquer SOC (Jesus; Moreira, 2018), auxiliando no processo de indexação e recuperação da informação, e, conseqüentemente, potencializando a qualidade da comunicação especializada (Sales; Café, 2009).

2.2 CATALOGAÇÃO E METADADOS

A Organização e Representação do Conhecimento e da Informação compreende, além da representação temática, a representação descritiva dos recursos informacionais. Este processo se completa na catalogação, em que é realizada a representação de um item e são definidos os seus pontos de acesso, para sua posterior recuperação, gerando um catálogo. Ao longo dos anos, o termo catalogação vinha sendo utilizado para os estudos e atividades apenas relacionados aos aspectos descritivos. Porém, atualmente a catalogação expressa o conceito de representação da informação, de forma que a representação temática é combinada à representação descritiva, juntamente com os dados de localização (Simões; Lima, 2020).

A representação temática já foi tratada nas seções anteriores, em que foi apresentado estudo de aspectos conceituais e semânticos de entidades em um domínio do conhecimento, além dos instrumentos desenvolvidos para classificação e indexação. Portanto nas próximas seções será abordada a catalogação, enquanto processo de representação informacional, porém com foco na representação descritiva, que, em sistemas informatizados, é formalizada por meio do uso de metadados. Como o objetivo desta pesquisa é elaborar uma proposta de modelo de representação para produtos do varejo, o foco para o estudo da catalogação serão os instrumentos utilizados para a representação descritiva, bem como seus princípios e formatos. Desta forma, esta seção está subdividida em duas subseções: Catalogação e Metadados.

2.2.1 Catalogação

A catalogação é um processo de representação de um recurso informacional, em que são considerados os aspectos descritivos e temáticos, propiciando a recuperação deste recurso por meio de pontos de acesso, utilizando mecanismos de busca em sistemas de recuperação da informação (Machado; Zafalon, 2020). O principal produto da catalogação é o catálogo, que reúne os registros bibliográficos, servindo de canal entre o acervo e seus usuários. Atua como um instrumento de comunicação e localização para acesso às obras:

Os catálogos viabilizam a busca, a recuperação e o acesso aos recursos informacionais para os usuários. Sendo que a construção dos catálogos é estabelecida pelos processos e técnicas da catalogação, que a partir de instrumentos como códigos, modelos e padrões reconhecidos

internacionalmente propiciam o tratamento dos recursos informacionais. Desse modo, a catalogação pode garantir a padronização, identificação, localização e a consistência dos dados bibliográficos, de autoridade e outros dados disponíveis no catálogo. (Machado; Arakaki, 2022, p. 2).

O processo da catalogação é composto por um conjunto de operações, utilizando códigos, normas e padrões, que compreendem o conteúdo e a forma de documentos originais, transformando em outro tipo de documento, de natureza instrumental e secundária, que serve como um meio para que o usuário acesse os documentos primários representados pelo processo de catalogação. O catálogo permite, portanto, a localização e acesso aos recursos informacionais (Machado; Zafalon, 2020), e as representações servem como meio para se chegar ao que se quer (Simões; Lima, 2020).

A representação descritiva é o processo de tratamento dos aspectos formais dos documentos. Os elementos descritivos distinguem, individualizam e identificam as particularidades de um recurso informacional, fazendo com que sua descrição seja distinta de seus similares (Machado; Zafalon, 2020). Por recurso informacional, entende-se: “[...] a informação objetivada no contexto de um campo do conhecimento podendo ser apresentado em uma estrutura analógica e/ou digital, com valor informacional que caracteriza a sua concepção intelectual expressa na corporificação de manifestações estruturadas na forma de itens.” (Santos; Simionato; Arakaki, 2014, p. 148). Os recursos informacionais apresentam características que são representadas por elementos descritivos específicos, que visam otimizar sua recuperação (Santos; Simionato; Arakaki, 2014). Estes elementos são formalizados em Sistemas de Recuperação da Informação por meio de metadados, que serão abordados na próxima subseção deste trabalho.

De acordo com Mey (1995), a catalogação tem como objetivo tornar o recurso informacional único, utilizando uma estrutura padronizada e sucinta dos dados e informações referentes a um recurso informacional. A padronização da descrição auxilia na construção de estruturas de representações completas e interoperáveis, fazendo com que os recursos informacionais sejam disponíveis, acessíveis, recuperáveis e únicos (Arakaki; Simionato; Santos, 2017).

Klim¹⁴ (1981, p. 55 *apud* Simões; Lima, 2020, p. 91) identifica três funções da representação que delineiam seu objetivo principal:

¹⁴ KLIM, I. L. Developing a theory of bibliographic description. **Scientific and Technical Information Processing**, [Russia], v. 1, p. 50-58, 1981.

- A função de informação, que consiste primeiramente em prover informação sobre a existência de um documento específico e, em segundo lugar, em comunicar diversos tipos de informação sobre o documento;
- A função de identificação, que consiste em assegurar que a representação e o documento a que ela corresponde sejam idênticos; daí a formulação prévia dos elementos de representação, com número suficiente de elementos que permitam sua identificação;
- A função de recuperação, que consiste no fato de que a representação é um meio de se recuperar um documento, qualquer que seja o sistema de recuperação da informação.

Porém, para um bom conjunto de elementos de representação, além de considerar as funções da representação elencadas acima, é importante também observar fatores que afetam a descrição, como o tipo de organização da informação, tipo e tamanho do sistema de recuperação da informação, fonte das descrições, demandas dos usuários etc. (Simões; Lima, 2020). Além disso, a construção da representação deve considerar que o recurso informacional pode estar fisicamente ou digitalmente em outro espaço, como também, no próprio meio em que a representação está disponível (Arakaki; Simionato; Santos, 2017). No caso do comércio eletrônico, a representação do produto em um catálogo é o meio pelo qual o consumidor pode pesquisar e obter informações sobre produtos específicos, que estão fisicamente nos estoques das lojas ou nos centros de distribuição.

Neste sentido, Simões e Lima (2020) destacam a importância de definir o escopo da catalogação de acordo com o contexto em que ocorre, contemplando a definição de objetivos e finalidades da catalogação naquele contexto, os recursos informacionais que serão representados, a definição do nível de detalhamento da descrição, a estrutura de metadados a ser utilizada, as tecnologias que serão adotadas, o controle de termos para representação dos conceitos identificados no documento ou objeto e a delimitação de responsabilidades dos agentes envolvidos. A relevância de cada um destes aspectos no processo de catalogação é delimitada pelos objetivos institucionais em questão (Simões; Lima, 2020).

A partir da década de 1960, com o desenvolvimento de recursos computacionais, a catalogação teve grandes avanços, marcados pelo desenvolvimento do Código de Catalogação Anglo-Americano (AACR2), pelo Padrão Internacional para Descrição Bibliográfica (ISBD) e pelo formato de intercâmbio de dados, denominado *Machine Readable Cataloging* (MARC), que promoveu o compartilhamento de registros bibliográficos legíveis por máquinas (Simões; Lima,

2020; Arakaki; Simionato; Santos, 2017). Estes instrumentos foram sendo revisados e atualizados, ao longo do século XX, para atender aos novos suportes de informação, aos novos recursos tecnológicos, que modernizaram os catálogos em bibliotecas, e às novas demandas por informação pelos usuários (Simões; Lima, 2020).

Também na década de 1960, foi realizada a *International Conference on Cataloguing Principles*, ocorrida em Paris, em 1961, que resultou na Declaração de Princípios Internacionais de Catalogação, ficando conhecida como Princípios de Paris. A Declaração de Princípios Internacionais de Catalogação oferece diretrizes para o desenvolvimento de códigos de catalogação (Arakaki; Simionato; Santos, 2017) e apresenta as principais funções dos catálogos (Machado; Zafalon, 2020). Segundo a Declaração, os catálogos devem permitir ao usuário:

[...] encontrar um único recurso ou um conjunto de recursos; identificar se a entidade descrita corresponde a entidade pesquisada, assim como deve ser possível distinguir entre entidades semelhantes; selecionar recursos que melhor satisfaçam as necessidades informacionais do usuário, descartando aqueles que não correspondam as mesmas; adquirir ou obter, promovendo as condições necessárias para garantir o acesso, a partir da compra, empréstimo, acesso online, entre outras possibilidades; navegar e explorar, a partir da organização lógica dos dados bibliográficos, dados de autoridade e relacionamentos entre entidades, em outros catálogos e fontes de informação. (International Federation of Library Associations, 2016, p. 11).

A partir da década de 1990, com o Seminário sobre Registros Bibliográficos, realizado em Estocolmo, iniciam-se os estudos de modelagem conceitual no universo bibliográfico, visando o desenvolvimento de novos modelos de organização da informação baseados nas novas tecnologias de informação e comunicação (Simões; Lima, 2020). Em 1998, foi apresentado o primeiro modelo conceitual, o *Functional Requirements for Bibliographic Records* (FRBR) (Simões; Lima, 2020). O FRBR é um modelo conceitual, abstrato, composto por entidades, atributos e relacionamentos, com a finalidade de orientar a construção de padrões, e não ser ele mesmo um padrão de catalogação (Machado; Zafalon, 2020). Machado e Zafalon (2020, p. 54) declaram que:

Os modelos conceituais no universo bibliográfico foram propostos para servir de orientação à catalogação e colocam-se como posições teóricas no sentido de orientar a construção de padrões, normas, códigos que direcionam os registros bibliográficos usando de uniformidade, a despeito das diferenças culturais, linguísticas, tipos de conteúdo e sobre os seus suportes informacionais.

O FRBR propõe dois objetivos: o primeiro indica a necessidade de relacionar as necessidades dos usuários com os dados bibliográficos, e o segundo trata da

relação de atributos e entidades com as tarefas dos usuários (International Federation of Library Associations And Institutions, 1997). Para atingir estes objetivos, o FRBR define três grupos de entidades com a forma de relacionamento entre eles: 1) apresentar produtos do trabalho intelectual ou artístico; 2) indicar responsabilidades pelos produtos do grupo 1; e 3) indicar assuntos das obras do grupo 1 (Machado; Zafalon, 2020). Os produtos intelectuais ou artísticos apresentados no grupo 1, são representados pelas quatro entidades abaixo (Machado; Zafalon, 2020):

- a) obra: é uma criação intelectual, uma entidade abstrata;
- b) expressão: é a realização intelectual de uma obra, é a forma pela qual a obra é realizada;
- c) manifestação: trata-se da materialização física de uma expressão de uma obra;
- d) item: exemplar único de uma manifestação, uma entidade concreta.

O FRBR apresenta quatro tarefas básicas que são desempenhadas pelos usuários na busca por informação (Machado; Zafalon, 2020):

- a) encontrar entidades que correspondam aos critérios de busca formulados;
- b) identificar uma entidade, ou seja, confirmar que a entidade descrita corresponde à entidade procurada;
- c) selecionar uma entidade que atenda aos requisitos do usuário, ou recusar uma entidade que não atenda;
- d) adquirir ou obter acesso à entidade selecionada.

A partir de 2010, foram realizados estudos para consolidação dos modelos conceituais para bibliotecas. Em 2016, o novo modelo, inicialmente denominado FRBR, passa a ser identificado como Modelo de Referência Bibliotecária da IFLA (IFLA LRM), sendo aprovado e publicado em 2017. Estes modelos serviram de base para o desenvolvimento de uma nova norma de catalogação, a *Resource Description and Access* (RDA), que é mais adequada para a representação da informação no ambiente digital (SIMÕES; LIMA, 2020).

A RDA foi desenvolvida pelo *Joint Steering Committee for Development of RDA*, com objetivo de substituir o AACR2, código de catalogação amplamente utilizado para criação de registros bibliográficos, e que é compatível com o formato MARC. A RDA se propõe a fornecer uma estrutura flexível que contemple a descrição tanto de recursos informacionais físicos quanto digitais, considerando as tarefas dos catálogos

propostas aos usuários (Arakaki; Simionato; Santos, 2017). Portanto, a estrutura da RDA é fundamentada nos elementos conceituais da FRBR, além de também estar pautada na Declaração de Princípios Internacionais de Catalogação.

Neste cenário de adequação dos instrumentos de catalogação às novas tecnologias e necessidades trazidas por elas, o processo de catalogação busca integrar todas as atividades do ciclo de representação da informação, desde o planejamento da estrutura do sistema, passando pela definição da infraestrutura, definição de metadados, chegando até a busca e recuperação da informação por meio da interface do sistema (Arakaki; Simionato; Santos, 2017). Desta forma, torna-se importante a definição de um padrão de metadados a ser utilizado para a formalização dos elementos descritivos e temáticos dos recursos informacionais, representados em sistemas informatizados.

2.2.2 Metadados

O processo de catalogação se complementa e se consolida com o uso de metadados (Simões; Lima, 2020), com a finalidade de representar a forma e o conteúdo de recursos informacionais em sistemas de informação (Santos; Simionato; Arakaki, 2014). Metadado etimologicamente significa “dado sobre dado” (National Information Standards Organization, 2010), termo cunhado na década de 1960, por Jack E. Myers, presidente e fundador da *The Metadata Corporation*, para descrever um conjunto de dados que podem ser utilizados para organização, representação e localização de recursos informacionais (Pomerantz, 2015; Smiraglia, 2005). Na perspectiva da CI, o conceito de metadado é atribuído a uma informação estruturada para ações de identificação, descoberta, seleção, uso, acesso e gerenciamento de dados, conjuntos de dados e outros recursos digitais (Arakaki; Arakaki, 2020; Pavão *et al.*, 2015). Os metadados tornam explícitos os diferentes aspectos de um recurso, como sua estrutura, conteúdo, qualidade, contexto, origem, propriedade e condição (Santos; Simionato; Arakaki, 2014; Machado; Arakaki, 2022).

No ambiente digital, os metadados são utilizados como elementos para busca e acesso aos dados descritivos de recursos na *web*. Para Alvarenga (2001), o metadado pode ser considerado como equivalente a ponto de acesso, termo da área de catalogação. Da mesma forma que ocorre na catalogação, os metadados não se restringem a dados descritivos, mas também são utilizados para representar dados

relacionados ao conteúdo de recursos informacionais (Alvarenga, 2001). Os metadados tem como função a descrição de documentos, além de permitir a pesquisa e descoberta. Eles também podem servir para distinguir entre documentos semelhantes (Buckland, 2017). De acordo com Pavão *et al.* (2015, p. 104): “Os metadados são usados para definir permissões, direitos de acesso, compartilhamento, reutilização, redistribuição e políticas, bem como os requisitos técnicos para visualização, acesso ou preservação de objetos digitalizados ou concebidos originalmente em formato digital.”

Os dados estruturados por meio da utilização de metadados desempenham papel importante em buscas em aplicações na *web*. Os dados estruturados na *web* trouxeram novas soluções para otimização de mecanismos de busca e integração de diferentes fontes de dados. Kejriwal *et al.* (2021) destacam que as áreas de comércio eletrônico e anúncios na *web* foram muito beneficiadas com o uso de dados estruturados, pois desta forma é mais fácil anunciar e encontrar produtos na *web*. Para melhorar a satisfação dos usuários na recuperação de informação por meio de sistemas automatizados, é importante que os metadados sejam completos, consistentes e que apresentem qualidade. Além disso, a organização da informação disposta em uma interface simples e amigável são fatores determinantes para o bom resultado da busca pelo usuário (Pavão *et al.*, 2015).

Um aspecto crítico, mas pouco explorado sobre metadados, é que eles podem ser portadores de informações sobre relacionamentos entre recursos informacionais. Este aspecto é importante na perspectiva da OC, pois esta área concentra-se fortemente no estabelecimento e representação de relacionamentos entre conceitos em um modelo conceitual, como os relacionamentos hierárquicos, associativos e de equivalência. Estes relacionamentos se manifestam como metadados em um sistema de informação e podem ser aproveitados para possibilitar a descoberta e compreensão de informações (Mayernik, 2020).

Arakaki (2019) categoriza os metadados de acordo com as tipologias abaixo:

- a) metadados administrativos: fornecem informações sobre origem e manutenção de um objeto, sendo usados para gerenciar e administrar coleções e recursos informacionais;
- b) metadados de autenticação: identificação, integridade e legitimidade de um recurso informacional;

- c) metadados de preservação: informações de preservação e conservação dos recursos;
- d) metadados de proveniência: dados sobre entidades, criação e modificações e seus relacionamentos;
- e) metadados técnicos: informações do sistema ou do recurso, relacionadas a como o sistema funciona;
- f) meta-metadados: informações sobre a criação de um registro ou de um conjunto de registros;
- g) metadados descritivos: características identificadoras e informações sobre os contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão;
- h) metadados de direitos: informações sobre propriedade e direitos autorais;
- i) metadados de acesso e uso: informações sobre como o recurso informacional foi acessado e utilizado;
- j) metadados estruturais: informações sobre a composição e a organização de um recurso informacional;
- k) linguagens de marcação: integram metadados e sinalizações para outros recursos estruturais ou semânticos.

Gilliland-Swetland (1999) categoriza os metadados de acordo com suas funções, organizando-os em cinco níveis: administrativos, para gerenciamento e administração dos recursos informacionais; descritivos, para descrição e identificação dos recursos; conservação; técnicos, sobre o funcionamento do sistema e comportamento dos metadados; e de uso, sobre tipo e nível de uso do recurso informacional.

O conjunto de vários elementos de metadados forma um esquema ou padrão de metadados que são compreendidos como coleções de propriedades pré-determinadas para descrever um recurso (National Information Standards Organization, 2010), garantindo a efetiva busca, acesso e recuperação de recursos no ambiente digital (Pavão *et al.*, 2015). Além disso, o uso de esquemas de metadados pode facilitar a interoperabilidade entre sistemas (National Information Standards Organization, 2010). O *Dublin Core Metadata Initiative* (DCMI) declara que: “Os esquemas são especificações processáveis por máquina que definem a estrutura e a sintaxe das especificações de metadados em uma linguagem de esquema formal.”

(Dublin Core, 2023, tradução nossa).¹⁵ Para Alves (2010, p. 47-48), padrões de metadados são:

[...] estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Estes conjuntos especificam regras semânticas para cada elemento de metadado, para o qual são atribuídos valores que, muitas vezes, podem ser oriundos de vocabulários controlados (National Information Standards Organization, 2010). Mayernik (2020) ressalta que, apesar de esquemas de metadados serem considerados dados estruturados, pois são de fato estruturas formadas por metadados, o nível de controle e padronização desta estrutura determina o quão estruturada ela é. Os elementos controlados podem exigir que as informações contidas no campo de metadado estejam em conformidade com uma sintaxe especificada, como a sintaxe para “dia-mês-ano”. Pode haver também campos de metadados em que esteja disponível um conjunto pré-determinado de valores para seu preenchimento, como por exemplo termos de um vocabulário controlado. Da mesma forma, podem haver campos de metadados sem controle de valores a serem preenchidos neste campo. Portanto, mesmo dentro de um esquema estruturado de metadados, pode haver metadados não estruturados (Mayernik, 2020).

Em relação ao uso de vocabulários controlados como instrumentos para estruturação de metadados, as taxonomias vêm sendo utilizadas para criação de metadados para descrição de objetos em ambientes digitais (Carlan; Medeiros, 2011), servindo para ordenamento e rotulação de metadados (Vital; Café, 2007). Neste sentido, há experiências de aplicação de taxonomias facetadas para organização da informação em ambientes digitais, possibilitando uma representação estruturada de objetos digitais, considerando tanto seus aspectos descritivos quanto semânticos e temáticos (Pontes; Lima, 2012).

A seleção e definição do padrão de metadados a ser utilizado em um sistema de informação para catalogação de recursos informacionais tem por finalidade escolher os elementos necessários para a descrição dos diferentes tipos de recursos que serão representados. De acordo com Santos, Simionato e Arakaki (2014, p. 152):

¹⁵ No original: “Schemas are machine-processable specifications which define the structure and syntax of metadata specifications in a formal schema language.” (Dublin Core, 2023).

Essa atividade tem como objetivo definir os elementos descritivos que farão parte do registro que representa cada item que compõe o acervo de uma determinada instituição e assim promover a gestão de acervos e de coleções pautada nos objetivos institucionais, no domínio do conhecimento e nas tarefas do usuário.

É importante destacar que nem todo atributo de um recurso informacional precisa ser descrito. Eles podem não ser considerados importantes no contexto em questão, ou podem não ser aplicáveis para um determinado recurso ou, ainda, podem não ser verificáveis (Hider, 2018).

A definição dos metadados deve ser uma ação consensual e levar em consideração a necessidade de interoperabilidade de seus dados com outros sistemas. No momento desta definição, os catalogadores devem considerar principalmente as necessidades dos usuários, além das demandas da instituição (Santos; Simionato; Arakaki, 2014). Neste processo de definição de metadados, pode-se optar por utilizar um único esquema de metadados ou um perfil de aplicação, em que podem ser selecionados e combinados elementos de diferentes esquemas em um esquema composto. Esta segunda opção pode suportar de forma mais adequada as exigências de uma aplicação em particular, sendo uma adaptação dos metadados para uma situação específica (Pavão *et al.*, 2015).

A construção de um padrão de metadados requer a adoção de procedimentos metodológicos para a definição dos metadados, em uma estrutura de descrição padronizada (Arakaki; Arakaki, 2020). De acordo com Coyle (2010), os metadados são construídos por seres humanos para um propósito, atividade ou problema específico, tendo como finalidade ser úteis de alguma forma. Portanto, os metadados são seletivos, pois são criados para propósitos específicos, e a sua criação ou seleção deve ser considerada como um tipo de ação situada em um contexto social. Desta forma, o estudo de metadados de algum sistema de informação pode revelar prioridades, expectativas e responsabilidades existentes em relação a estes sistemas (Mayernik, 2020).

Pavão *et al.* (2015) destaca importância de confrontar o esquema de metadados que está sendo selecionado ou desenvolvido, com os seis princípios dos “bons metadados” apresentados pela *National Information Standards Organization* (2007, p. 61-62, tradução nossa), sendo eles:

Princípio para metadados 1: Bons metadados estão em conformidade com os padrões da comunidade de uma forma apropriada para os materiais da coleção, usuários da coleção e usos futuros, atuais e potenciais da coleção.
Princípio para metadados 2: Bons metadados suportam a interoperabilidade.
Princípio para metadados 3: Bons metadados usam controle de autoridade e padrões de conteúdo para descrever objetos e dispor objetos relacionados.
Princípio para metadados 4: Bons metadados incluem uma declaração clara das condições e termos de uso do objeto digital.
Princípio para metadados 5: Bons metadados suportam a curadoria de longo prazo e a preservação de objetos em coleções.
Princípio para metadados 6: Bons registros de metadados são objetos em si e, portanto, devem ter as qualidades de bons objetos, incluindo autoridade, autenticidade, arquivabilidade, persistência e identificação única.¹⁶

Para a implementação dos metadados, é importante que seja realizada uma análise do domínio para que, dessa forma, sejam criadas estruturas de metadados que garantam a utilização e interpretação dos mesmos pelos seus usuários. Nesta perspectiva, o modelo conceitual FRBR pode ser usado como orientação para a definição dos elementos que irão compor a representação dos recursos informacionais (Santos; Simionato; Arakaki, 2014). Destaca-se que os instrumentos utilizados para definição dos metadados estão diretamente relacionados ao tipo de recurso informacional que se está representando, ao domínio do conhecimento em questão e às políticas institucionais. Por este motivo, é importante que em um projeto para adoção de metadados, sejam registradas as decisões do processo, os metadados definidos para a descrição dos diferentes tipos de recursos informacionais, o vocabulário controlado utilizado, os registros de autoridade e as diretrizes para o preenchimento dos valores dos metadados (Santos; Simionato; Arakaki, 2014).

Os metadados podem ser criados por profissionais especializados, como profissionais da informação ou por gerentes de bancos de dados. Podem contar também com a colaboração de especialistas do domínio do conhecimento específico, pois estes têm conhecimento sobre como os recursos relacionados à sua área são criados e usados, podendo fornecer informações importantes para a estruturação de metadados. Os metadados também podem ser gerados de forma automática tanto

¹⁶ No original: Metadata Principle 1: Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection. Metadata Principle 2: Good metadata supports interoperability. Metadata Principle 3: Good metadata uses authority control and content standards to describe objects and collocate related objects. Metadata Principle 4: Good metadata includes a clear statement of the conditions and terms of use for the digital object. Metadata Principle 5: Good metadata supports the long-term curation and preservation of objects in collections. Metadata Principle 6: Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification. (National Information Standards Organization, 2007, p. 61-62).

para criação de metadados técnicos ou de proveniência, por exemplo, que são inerentes a um sistema digital, quanto para metadados descritivos. Neste caso, duas técnicas automatizadas podem ser utilizadas para a criação de metadados (Greenberg, 2004): extração ou coleta. Na extração, um algoritmo extrai automaticamente metadados do conteúdo de um recurso. Na coleta, metadados de recursos distribuídos são compilados automaticamente.

Concluindo, há diversos padrões de metadados existentes, podendo apresentar abrangência mais geral ou especializada em um domínio de conhecimento ou por tipos de recursos específicos. Esta pesquisa tem o objetivo de propor um modelo de representação para produtos do varejo, que possa ser adaptado e aplicado em diferentes contextos e sistemas. Portanto, o foco neste trabalho serão os metadados descritivos, visando identificar e organizar os elementos que descrevem e representam o produto.

3 METODOLOGIA E PROCEDIMENTOS

O método pode ser compreendido como um conjunto de procedimentos, regras e técnicas adotado para realização de uma pesquisa científica. A escolha do método varia de acordo com a natureza da pesquisa (Matias-Pereira, 2016). “A metodologia é o emprego do conjunto dos métodos, procedimentos e técnicas que cada ciência em particular põe em ação para alcançar os seus objetivos.” (Matias-Pereira, 2016, p. 43).

Nas próximas seções, será apresentada a contextualização e a caracterização da pesquisa, além dos procedimentos e técnicas que serão empregados.

3.1 CONTEXTUALIZAÇÃO DA PESQUISA

Os dados do corpus de pesquisa foram coletados em um contexto real, sendo este a instituição em que a pesquisadora atua profissionalmente, devido à facilidade de acesso aos dados e familiaridade com o domínio de conhecimento. São dados referentes a produtos do varejo alimentar comercializados entre empresas (B2B), de indústrias para minimercados localizados em São Paulo e região metropolitana, com perspectivas de expansão, através de plataforma de comércio eletrônico do ecossistema Hindiana.

O fundo de investimentos Hindiana sustenta um ecossistema formado por diversas empresas que atuam conjuntamente com o objetivo de conectar pequenos varejistas (minimercados) a grandes indústrias, promovendo mais eficiência, redução de custos, maior acesso a ofertas, melhor adequação de preços e de sortimento de produtos (Fernandes; Lopes; Zuzunaga, 2023).

O ecossistema Hindiana é composto por quatro empresas, cada uma focada em sanar diferentes dores dos elos da cadeia de valor (Hindiana, 2022). As quatro empresas são:

- a) Izio: oferece campanhas de marketing e soluções para cada grupo de consumidores, através de uma plataforma que conecta a indústria, o varejo e o consumidor;
- b) Yandeh: plataforma de marketplace phigital (digital e física) que oferece serviços de abastecimento, logística e gestão para o varejo, através da comercialização de produtos na plataforma, conectando fornecedores, vendedores, promotores e varejistas;

- c) Alpe: proporciona soluções financeiras para pagamento e crédito;
- d) Girotrade: oferece serviço fiscal e de logística para toda a cadeia de abastecimento.

A Yandeh é responsável pelo cadastro dos produtos que são comercializados na plataforma, além de fornecer cadastro de produtos para sistemas de outras empresas do ecossistema. Os dados dos produtos cadastrados também servem de insumo para recursos de análises de dados e algoritmos de inteligência artificial, proporcionando inteligência comercial para o varejista, como recomendações de compra, gestão de estoque, campanhas de marketing etc.

O cadastro de produtos é realizado por operadores de cadastro em um sistema informatizado, com funcionalidades para registrar e gerenciar os seguintes dados:

- a) marcas e fabricantes;
- b) tipos de produto (chamado de “linha” no sistema): designa o que é o produto, por exemplo, biscoito, geleia, goma de mascar, entre outros;
- c) taxonomia: organiza os produtos por meio de seu agrupamento;
- d) características dos produtos, divididas entre:
 - primárias: identificam e diferenciam o produto, compõem a descrição do produto que aparece na plataforma de comércio eletrônico;
 - secundárias: prazo de validade do produto, chamado de *shelf life*;
 - logísticas: peso líquido, peso bruto e peso drenado; e
 - fiscais: código NCM (Nomenclatura Comum do Mercosul);
- e) imagem do produto: foto da embalagem do produto;
- f) códigos das embalagens dos produtos: código único que identifica o registro do produto, chamado de GTIN (*Global Trade Item Number*). O GTIN é o número correspondente ao código de barras, que é atribuído ao produto pelo GS1.

Atualmente, a taxonomia está estruturada em 7 departamentos (primeiro nível), 19 seções (segundo nível), 106 categorias (terceiro nível), 483 subcategorias (quarto nível) e 2.062 segmentos (quinto e último nível). Para cada segmento, são relacionadas linhas de produtos, consideradas o tipo de produto. São 2.753 linhas de produtos. Para cada linha de produto, há uma definição e uma padronização de quais características são associadas a cada linha, como as características cor, sabor, tamanho etc.

Ao cadastrar o produto, o operador registra no sistema a linha de produto, marca e fabricante, taxonomia, características e preenche cada característica com os valores referentes ao produto em questão, a partir dos dados recebidos em planilhas com descrições dos produtos, em consulta em repositório de informações sobre produtos e em pesquisa na *internet* em fontes como:

- a) *Verified by GS1*: <https://www.gs1.org/services/verified-by-gs1>. Utilizado para consulta de fabricantes a partir do GTIN do produto;
- b) Cosmos Bluesoft: <https://cosmos.bluesoft.com.br/>. Utilizado para consulta de informações sobre o produto, como descrição e unidades de embalagens, a partir do GTIN do produto;
- c) Portal da Nota Fiscal Eletrônica – SVRS: <https://dfe-portal.svrs.rs.gov.br/Nfe/Gtin>. Utilizado para consulta de informações sobre o produto, como descrição, a partir do GTIN do produto.

A partir das informações cadastradas, o sistema gera a descrição do produto, ou seja, o nome do produto que aparecerá na plataforma de comércio eletrônico.

O público que acessa a plataforma de comércio eletrônico é composto por varejistas de minimercados, que fazem parte do varejo não especializado, ou seja, são estabelecimentos comerciais com venda predominante de produtos alimentícios e artigos de primeira necessidade, com pouca variedade. Portanto, para este público, o nível de detalhamento das informações dos produtos é baixo, comparado a lojas especializadas. Ou seja, assim como nas bibliotecas, em que em uma biblioteca especializada a representação do recurso informacional necessita de uma especificação maior do que em uma biblioteca pública, o cadastro de um produto para uma loja especializada deve ser mais detalhado do que para um minimercado, por exemplo.

Quanto ao público interno, que realiza o cadastro dos produtos na plataforma, este é composto por operadores de cadastro que, apesar de possuírem experiência no cadastro de produtos, não possuem especialização formal no domínio de conhecimento de produtos do varejo, e também não possuem conhecimento formal sobre os sistemas e processos para organização e representação da informação.

Com base nas informações apresentadas sobre o contexto, considerando também o problema e objetivos da pesquisa, é possível delinear algumas condições que, juntamente com a delimitação do domínio, podem guiar o desenvolvimento do modelo de representação desta pesquisa. Portanto, o modelo desenvolvido deve:

- a) ser compreendido e utilizado por humanos, visto que o cadastro é realizado por operadores;
- b) ser compatível com sistemas de informações, visto que o operador registra as informações no sistema;
- c) prever a utilização do modelo por algoritmos de inteligência artificial para classificação e recomendação;
- d) servir como base para navegação em plataformas de comércio eletrônico;
- e) o nível de detalhamento da representação não precisa ser alto, pois será utilizado por público não especializado;
- f) considerar variações terminológicas, devido à perspectiva de expansão da plataforma para além do Estado de São Paulo;
- g) a organização e representação deve atender a necessidade de compra e consumo do usuário.

Os pontos acima elencados, podem ser considerados no desenvolvimento de modelos de representação em outros contextos de pesquisa e aplicação, ajustando de acordo com a realidade e necessidade em questão.

3.2 CARACTERIZAÇÃO DA PESQUISA

Para o desenvolvimento desta pesquisa foi utilizada abordagem qualitativa, buscando compreender os elementos que constituem o objeto estudado, visando seu aprofundamento por meio de interpretação e estruturação dos dados coletados (Michel, 2015). De acordo com Gerhardt e Silveira (2009), a pesquisa qualitativa não se preocupa com a representatividade quantitativa, mas sim com o aprofundamento da compreensão de um fenômeno. “Na pesquisa qualitativa, a verdade não se comprova numérica ou estatisticamente; ela surge na experimentação empírica, a partir de análise feita de forma detalhada, abrangente, consistente e coerente, e na argumentação lógica das ideias.” (Michel, 2015, p. 40).

A pesquisa pode ser classificada de acordo com seus fins (resultado) e seus meios (processo). Quanto aos seus fins, foi realizada uma pesquisa básica, pois visa compreender as características relacionadas ao objeto de estudo, por meio de revisão de literatura e análise documental; além de uma pesquisa aplicada (Michel, 2015), visto que a intenção deste trabalho é gerar como resultado um modelo de

representação que possa ser reproduzido e aplicado, revelando o aspecto utilitário da pesquisa.

Quanto aos meios, esta pesquisa é classificada, em sua fase preliminar, como um estudo exploratório, buscando “[...] explicar um problema a partir de referências teóricas publicadas em documentos, dispensando a elaboração de hipóteses.” (Michel, 2015, p. 47). A pesquisa exploratória proporciona maior familiaridade com o tema apresentado e uma visão geral do objeto (Lise *et al.*, 2018). Por se tratar de uma pesquisa aplicada, pode ser classificada como um estudo de ação, pois é direcionada a resolver um problema concreto de mercado (Matias-Pereira, 2016).

3.3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa trata sobre produtos do varejo, mais especificamente do varejo alimentar, tendo como recorte a categoria de biscoitos. Portanto estes serão o domínio do conhecimento e o recorte para o qual será elaborado modelo de representação. Com objetivo de compreender e modelar este domínio, para sua representação, foi utilizada Análise de Domínio como metodologia para a elaboração de SOCs. A partir desta metodologia, é possível mapear e compreender domínios do conhecimento, a fim de desenvolver sistemas e processos de OC.

A utilização da AD como base metodológica, buscou uma melhor compreensão do domínio de conhecimento de produtos do varejo, por meio de revisão de literatura na área, configurando o corpus textual especializado da pesquisa, identificando os principais conceitos e como eles se relacionam no domínio de conhecimento. Este material serve de apoio na fase de análise de dados do corpus de pesquisa, com objetivo de validar os conceitos e relações identificadas nesta fase.

Em relação ao corpus de pesquisa, foi realizada análise qualitativa com os dados de biscoitos coletados de base de dados com descrições dos produtos cadastrados na plataforma de comércio eletrônico. Os procedimentos para a análise dos dados do corpus de pesquisa foram organizados nas seguintes etapas: planejamento e coleta de dados, preparação dos dados, tratamento dos dados, análise dos dados e apresentação dos resultados.

3.3.1 Planejamento e coleta de dados

Nesta etapa, é definido o escopo da pesquisa e sua viabilidade. Quanto ao escopo, foram coletados e analisados dados de um contexto real, no domínio de produtos do varejo, conforme relatado na seção 3.1.

Como recorte da pesquisa, foram considerados produtos de apenas uma categoria da taxonomia já existente na empresa. Foi selecionada a categoria de biscoitos para compor o recorte da pesquisa, pois para o contexto específico, os produtos de mercearia seca, como biscoitos, são priorizados no cadastro, devido a maior praticidade logística em comparação com produtos perecíveis, por exemplo. Atualmente na empresa em questão, a categoria de biscoitos representa 10,64% de produtos de mercearia seca cadastrados no sistema, ficando atrás apenas das categorias de doces, sobremesas e preparos, e bomboniere, com 12,45% e 10,78% respectivamente. Estas duas categorias abrangem uma variação maior de tipos diferentes de produtos, e por isso não foram selecionadas para esta pesquisa. Já na categoria de biscoitos, pode-se encontrar 22 segmentos diferentes, mas todos considerados como tipos de biscoitos. Além disso, o Brasil movimentou R\$ 29 bilhões em 2022 em valor e 1,5 milhões de toneladas em volume de vendas de biscoitos. Pesquisa realizada pela Indexsa, revelou que o biscoito é um alimento que tem penetração em 99,7% nos lares brasileiros. Estes números demonstram a importância desta categoria de produtos alimentícios no mercado brasileiro e nos hábitos de consumo dos brasileiros (Biscoito [...], 2023).

O corpus de pesquisa é formado por base de dados com descrições completas dos produtos cadastrados na plataforma de comércio eletrônico em questão, compondo o universo desta pesquisa. A amostra de dados é proveniente de uma relação de itens constantes na base de dados do cadastro, no sistema de cadastro, já tratados pela empresa, contendo descrição, marca e fabricante, linha, taxonomia, características primárias, secundárias, logísticas, fiscais e códigos de embalagens.

A amostra de pesquisa é composta por um recorte quantitativo de produtos que estão categorizados como biscoitos na relação de produtos citada acima, com saneamento alto, ou seja, as informações já foram validadas pela equipe de cadastro da empresa. Dentro da categoria de biscoitos, há 22 segmentos (último nível da taxonomia). Dentro de cada segmento há uma quantidade de produtos que estão cadastrados no sistema naquele segmento, conforme abaixo:

Quadro 1 – Quantificação de produtos por segmento

Segmentos	Quantidade	Percentual
Biscoito Recheado	402	20,68%
Biscoito Doce Saborizado	213	10,96%
Biscoito Wafer	213	10,96%
Biscoito Cookie	144	7,41%
Biscoito Amanteigado e Sequilho	131	6,74%
Biscoito Salgado Snacks e Porcionado	131	6,74%
Biscoito Cream Cracker	130	6,69%
Biscoito de Polvilho	122	6,28%
Biscoito Rosquinha	107	5,50%
Biscoito Maisena	69	3,55%
Biscoito Salgado Saborizado	51	2,62%
Biscoito Doce Integral	49	2,52%
Biscoito Água e Sal	42	2,16%
Biscoito Maria	40	2,06%
Biscoito com Cobertura	34	1,75%
Biscoito Doce de Arroz	22	1,13%
Biscoito Champanhe	14	0,72%
Biscoito Salgado de Arroz	11	0,57%
Biscoito Água	6	0,31%
Biscoito Recheado e com Cobertura	6	0,31%
Biscoito Cristais de Açúcar	4	0,21%
Biscoito Sabor Manteiga	3	0,15%
Total	1944	100,00%

Fonte: Elaborado pela autora.

Como amostra desta pesquisa, foi selecionado um produto de cada segmento, pelo critério de maior extensão de sua descrição, compreendendo que quanto maior a descrição, mais elementos descritivos este produto apresenta para análise. Com o objetivo de favorecer a replicabilidade e a reuso dos dados obtidos nesta pesquisa, a amostra do corpus de pesquisa composta por um total de 22 produtos está disponível no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11081212> (APÊNDICE A - Descrições dos produtos).

Quanto à viabilidade da pesquisa, foi concedida autorização pela empresa para a pesquisadora, para acesso e utilização dos dados citados acima nesta pesquisa.

3.3.2 Preparação dos dados

Nesta etapa, os dados coletados são organizados em planilha, utilizando *software* Excel, para seu posterior tratamento e análise. Os dados compilados e

organizados foram analisados para verificar se há algum registro corrompido ou que impeça sua identificação, bem como registros duplicados. Não foram identificados dados corrompidos.

Todas as colunas dos dados foram nomeadas de forma que atendam ao objetivo da pesquisa e facilitem a compreensão e análise. Nesta fase, foi avaliado se há dados na base fornecida pela empresa que, mesmo estando corretos, não serão considerados para o corpus da pesquisa, apresentando a devida justificativa para sua remoção da amostra. No segmento “Biscoito Salgado Snacks e Porcionado”, o item com descrição mais extensa foi um produto com linha de produto Snack, e não Biscoito (SNACK FIT FOOD IMMUNE NUTS UVA-PASSA, AMENDOIM, CASTANHA DO PARÁ, CRANBERRY E NIBS DE CACAU PACOTE 30 G). Portanto este item foi desconsiderado e selecionado o item com segunda maior extensão da descrição (BISCOITO SALGADO CLUB SOCIAL ORIGINAL PACOTE 12 UN 288 G EMBALAGEM ECONÔMICA LEVE MAIS PAGUE MENOS).

3.3.3 Tratamento dos dados

Os dados do corpus de pesquisa coletados no sistema utilizado para cadastro de produtos já foram tratados pela empresa, contendo informação de descrição, marca e fabricante, linha, taxonomia, características primárias, secundárias, logísticas, fiscais e códigos de embalagens, de forma estruturada. Como estes dados já foram tratados e validados na empresa, não foi realizada pesquisa adicional, para buscar por mais informações sobre estes produtos. Portanto a análise foi realizada sobre os dados já estruturados destes produtos, priorizando os dados constantes na descrição do produto, como marca, linha e características primárias, visto que o escopo deste trabalho se concentra nos aspectos descritivos dos produtos.

3.3.4 Análise dos dados

A partir da listagem de descrições coletadas e tratadas, inicia-se a análise dos dados, de acordo com as etapas abaixo:

- a) extrair termos das descrições dos produtos;
- b) identificar as características dos conceitos selecionados, designados pelos termos, por meio da elaboração de definições;

- c) agrupar os conceitos em categorias com base nas características identificadas;
- d) relacionar as categorias, identificando as relações, inclusive entre classes e suas características;
- e) identificar variações denominativas dos conceitos e definir o termo autorizado;
- f) estruturar padrão de metadados para representar formalmente a estrutura conceitual elaborada;
- g) validar a estrutura perante os conceitos e terminologia identificados no domínio de produtos do varejo, para fins de garantia literária.

As etapas listadas acima foram fundamentadas nas metodologias utilizadas para o desenvolvimento de SOCs, apresentadas nas seções de referencial teórico. Estes procedimentos visam atender aos objetivos específicos definidos para esta pesquisa, conforme quadro abaixo:

Quadro 2 – Procedimentos e Objetivos específicos

Procedimentos	Objetivos específicos
Extrair termos das descrições dos produtos	Identificar e analisar os atributos relevantes dos produtos que compõem o recorte da pesquisa
Identificar as características dos conceitos selecionados, designados pelos termos, por meio da elaboração de definições	Identificar e analisar os atributos relevantes dos produtos que compõem o recorte da pesquisa
Agrupar os conceitos em categorias com base nas características identificadas	Identificar e relacionar categorias de representação a partir destes atributos
Relacionar as categorias, identificando as relações, inclusive entre classes e suas características	Identificar e relacionar categorias de representação a partir destes atributos
Identificar variações denominativas dos conceitos e definir o termo autorizado	Estabelecer controle terminológico dos termos adotados nas categorias
Estruturar padrão de metadados para representar formalmente a estrutura conceitual elaborada	Estruturar modelo de representação para o recorte estudado

Validar a estrutura perante os conceitos e terminologia identificados no domínio de produtos do varejo, para fins de garantia literária	Estruturar modelo de representação para o recorte estudado
---	--

Fonte: Elaborado pela autora.

3.3.5 Apresentação dos resultados

Por fim, será apresentado modelo de representação de produtos do varejo, em uma estrutura organizada de maneira sistemática e formal, acompanhado de uma síntese dos procedimentos empregados em seu desenvolvimento, com o intuito de viabilizar sua adoção e adaptação em diferentes contextos.

3.4 ANÁLISE DE DOMÍNIO

A AD caracteriza-se como uma nova abordagem metodológica para a CI, especialmente para a área da Organização do Conhecimento. A partir desta metodologia, é possível mapear e compreender domínios do conhecimento, a fim de desenvolver sistemas e processos de OC.

A AD foi iniciada na CI por meio de trabalho publicado por Hjørland e Albrechtsen (1995) e revigorado pelo trabalho de Hjørland (1998; 2002) e Tennis (2012). Para López-Huertas (2015, p. 570, tradução nossa), “[...] Análise de Domínio foi concebida para descobrir as estruturas de conhecimento, dinâmicas, linguagem, padrões de comunicação e comportamento de cooperação de domínios especializados.”¹⁷

A AD surge no contexto da informação especializada em CI e da reflexão sobre a competência do profissional da informação, que é qualificado para tratar da infraestrutura e processos de informação, mas não é um especialista sobre assunto ao qual a informação se refere (Hjørland, 2002).

A AD tem como foco a importância do conhecimento sobre o assunto, por ser fundamental para a comunicação do conhecimento (Hjørland, 2017). A CI trata da

¹⁷ No original: “Domain analysis was conceived for finding out the knowledge structures, dynamics, language and communication patterns and cooperation behavior of specialized domains.” (López-Huertas, 2015, p. 570).

mediação da informação e conhecimento, que se referem a algum assunto específico. Portanto, mediar a informação requer algum nível de conhecimento sobre o assunto, dependendo do contexto. Por exemplo, o profissional da informação deve possuir um conhecimento mais aprofundado do assunto de uma biblioteca especializada, do que teria em uma biblioteca pública, por abranger assuntos mais gerais.

Neste sentido, a AD é uma metodologia que pode ser usada pelos profissionais da informação para otimizar sistemas e serviços de informação, que são sua especialidade, a partir da perspectiva do assunto específico que está sendo tratado, e não de forma generalista. Portanto, através da AD o profissional da informação teria condições de compreender o assunto do domínio específico de conhecimento, a partir da terminologia especializada da área.

Hjørland (2017) apresenta algumas variedades da AD. Em relação à CI, a AD pode ser compreendida em um sentido amplo ou estrito. Em sentido amplo, a AD inclui estudos bibliométricos, análise de facetas, ou ainda qualquer uma das 11 abordagens apresentadas abaixo. Em sentido estrito, estudos serão considerados como AD apenas se considerarem diferentes teorias, paradigmas ou tradições no domínio.

Hjørland (2002) sugere 11 abordagens pelas quais a CI pode tratar de um determinado domínio, através da AD. Ele recomenda que estas abordagens sejam utilizadas de forma combinada e destaca a importância do conhecimento destas abordagens por profissionais da informação.

1. Produção e avaliação de guias de literatura e portais de assuntos;
2. Produção e avaliação de sistemas de classificação e tesouros;
3. Pesquisa sobre competências em indexação e recuperação de informação em especialidades;
4. Conhecimento de estudos empíricos de usuários em áreas temáticas;
5. Produção e interpretação de estudos bibliométricos;
6. Estudos históricos de estruturas e serviços de informação em domínios;
7. Estudos de documentos em domínios do conhecimento;
8. Estudos epistemológicos e críticos de diferentes paradigmas, pressupostos e interesses em domínios;
9. Conhecimentos de estudos terminológicos, linguagens para fins especiais e análise do discurso em áreas do conhecimento;
10. Estudos de estruturas e instituições de comunicação científica e profissional em um domínio;
11. Conhecimento de métodos e resultados de estudos analíticos de domínio sobre cognição profissional, representação do conhecimento em ciência da computação e inteligência artificial (HJØRLAND, 2017, p. 2, tradução nossa).¹⁸

¹⁸ No original: "1) Production and evaluation of literature guides and subject gateways; 2) Production and evaluation of special classifications and thesauri; 3) Research on competencies in indexing and retrieval of information in specialties; 4) Knowledge of empirical user studies in subject areas; 5) Production and interpretation of bibliometric studies; 6) Historical studies of information structures

Smiraglia (2015) propõe a exclusão das abordagens 3 e 10, e a inclusão de semântica de base de dados e análise do discurso. Guimarães e Tognoli (2015) também sugerem a inclusão de uma nova abordagem, sobre proveniência.

Dentre as abordagens citadas, as abordagens destacadas abaixo apresentam relação direta com a área da OC. Além disso, podemos identificar relação também com a abordagem sobre a (11) representação do conhecimento em Ciência da Computação e Inteligência Artificial:

- a) (2) a produção e avaliação de sistemas de classificação e tesouros;
- b) (3) competências em indexação e recuperação da informação;
- c) (9) estudos terminológicos, linguagens para fins especiais e análise do discurso em áreas do conhecimento.

Quanto a produção e avaliação de sistemas de classificação e tesouros, Hjørland (2002) destaca que para compreender, desenvolver e avaliar esquemas de classificação é necessário possuir conhecimento sobre os pressupostos epistemológicos da área. Hjørland (2002) também critica os métodos de construção de sistemas de classificação, por serem muito empíricos, sem o rigor da metodologia científica. Neste sentido, a AD pode auxiliar como base metodológica para construção de SOCs. Nesta pesquisa, será elaborado um modelo de representação de produtos do varejo, que será construído com base nas estruturas de SOCs. Portanto esta abordagem da AD se aplica a esta pesquisa.

Em relação às competências em indexação e recuperação da informação, Hjørland (2002, p. 430, tradução nossa) destaca as diferentes demandas de informação em sistemas, afirmando que:

A indexação e a recuperação da informação são sempre específicas. A CI tradicional tem, no entanto, ignorado em grande parte a forma como diferentes domínios podem colocar demandas diferentes em sistemas para organizar e recuperar documentos. Um foco mais forte em diferentes domínios pode tornar nosso campo mais realista e nossos mestres mais relevantes em diferentes ambientes.¹⁹

and services in domains; 7) Studies of documents and genres in knowledge domains; 8) Epistemological and critical studies of different paradigms, assumptions, and interests in domains; 9) Knowledge of terminological studies, LSP (languages for special purposes), and discourse analysis in knowledge fields; 10) Studies of structures and institutions in scientific and professional communication in a domain; 11) Knowledge of methods and results from domain-analytic studies on professional cognition, knowledge representation in computer science, and artificial intelligence.” (Hjørland, 2017, p. 2).

¹⁹ No original: “Indexing and retrieving information is always specific. Main stream IS has, however, largely ignored the way different domains may put different demands on systems for organising and

Esta abordagem também será adotada nesta pesquisa, porque será necessário considerar a realidade dos usuários do sistema, tanto os usuários que pesquisam e acessam as representações dos produtos para compra, quanto os usuários internos da instituição, que atuam como indexadores deste conteúdo e não são especializados nem no domínio do conhecimento em questão nem em organização e representação da informação.

Quanto aos estudos terminológicos, Hjørland (2002) declara que o domínio apresenta uma terminologia própria e as linguagens especializadas refletem o conhecimento especializado daquele domínio. Segundo Hjørland (2002, p. 446, tradução nossa): “Linguagem e terminologia são objetos muito importantes para CI porque afetam nosso pensamento e, portanto, as perguntas que colocamos em bancos de dados, bem como os textos que pesquisamos”.²⁰ Esta abordagem também será utilizada nesta pesquisa, visto que os produtos do varejo configuram um domínio em que há uma terminologia própria, incluindo regionalismos e jargões na área, sendo importante a perspectiva da terminologia para a definição dos termos que serão adotados no modelo de representação de produtos do varejo.

Em relação a representação do conhecimento em Ciência da Computação e Inteligência Artificial, Hjørland (2002) declara que a AD é usada como um método para desenvolvimento de engenharia de *software*, que busca capturar os pontos em comum e as variações de sistemas dentro de um domínio, aumentando eficiência e reusabilidade. Neste sentido, esta pesquisa não pretende se aprofundar nos aspectos técnicos das aplicações tecnológicas para implementação do modelo de representação proposto. Porém, considera o contexto do comércio eletrônico, pois é neste ambiente em que ocorre a representação do produto. Portanto, reconhece-se nesta pesquisa, que o desenvolvimento de um modelo de representação de produto do varejo servirá para aplicações de tecnologia, como plataformas digitais, além de algoritmos de inteligência artificial, por exemplo, para classificação automatizada, recomendação de produtos e mecanismos de busca.

retrieving documents. A stronger focus on different domains may make our field more realistic and our masters more relevant in different environments.” (Hjørland, 2002, p. 430).

²⁰ No original: “Language and terminology are very important objects for IS because they affect our thinking and thus the questions we put to databases as well as the text we search.” (Hjørland, 2002, p. 446).

A partir destas abordagens, observa-se que para modelar o conhecimento de uma área especializada, para o desenvolvimento de sistemas e processos de OC, para organização e recuperação da informação, é necessário conhecer mais profundamente os assuntos de que tratam a área, sendo a AD uma abordagem metodológica propícia para o estudo de um domínio específico do conhecimento.

Na área da OC, podemos destacar mais especificamente a representação do conhecimento como forma de materializar o trabalho de modelagem conceitual de um domínio do conhecimento, através de uma linguagem especializada ou sistema de classificação que pode ser compreendido por humanos ou máquinas.

Para Barros e Laipelt (2021, p. 442):

[...] a representação do conhecimento estará limitada a um determinado contexto e objetivo, representando, portanto, apenas uma parte da realidade, ou seja, um dado domínio de conhecimento. A ORC busca, então, trabalhar de forma aprofundada com a elaboração de sistemas, visando à representação de um determinado domínio.

Considerando a utilização da AD como metodologia para construção de SOCs, Hjørland (2016b) destaca que um mesmo objeto de estudo pode ser classificado sob várias perspectivas. Portanto, é importante a análise da epistemologia do domínio para compreender quais conceitos são centrais e podem representar algum consenso na área. Dependendo da abordagem da análise, pode ser importante trazer várias perspectivas do domínio para o SOC que está sendo construído.

3.4.1 Definição de Domínio

A palavra “domínio” pode ter vários sentidos de acordo com o contexto. Para a AD, pode-se entender domínio como sendo uma disciplina do conhecimento, uma especialização socialmente institucionalizada na divisão do trabalho, uma comunidade discursiva, ou até mesmo um hobby. Hjørland (2017, p. 6, tradução nossa) afirma que: “Um domínio é um corpo de conhecimento, definido social e teoricamente como o conhecimento de um grupo de pessoas que compartilham compromissos ontológicos e epistemológicos.”²¹

Mai (2005) e Shapere (1984) acreditam que o domínio seja um conceito que está aberto e evolui conforme seja aplicado na pesquisa e na prática. Mai (2005, p.

²¹ No original: “A domain is a body of knowledge, defined socially and theoretically as the knowledge of a group of people sharing ontological and epistemological commitments.” (Hjørland, 2017, p. 6).

605, tradução nossa) afirma que o “Domínio é um conceito aberto e em evolução que se desenvolverá à medida que o conceito seja usado e aplicado na pesquisa e na prática.”²² Ou seja, o domínio será determinado de acordo com o objeto de análise e seu recorte. Portanto, pode ser amplo como uma disciplina do conhecimento ou mais especializado, como uma subárea desta disciplina, ou até mesmo um campo interdisciplinar. Esta visão vai ao encontro de Dam Christensen²³ (2007 *apud* Hjørland, 2017) que acredita que o domínio sempre é ambíguo e está mais relacionado com processos do que com um objeto inteiro e estável, dependendo mais do recorte da análise.

Para Sugimoto e Weingart (2015), a declaração de que o domínio está relacionado com comunidades de discurso leva a ideia de que se trata de disciplinas do conhecimento, pois esta tem sido uma característica reconhecida da disciplinaridade. Porém, como já citado, o domínio não se limitaria apenas a disciplinas já consolidadas. Abrangeria também uma especialização socialmente institucionalizada na divisão do trabalho e até grupos com um *hobby* específico. Neste sentido, pensando nesta diversidade, especialmente no meio científico, pode-se estudar também a aplicação da AD em campos de conhecimento interdisciplinar.

Buscando compreender e delimitar um domínio em contextos interdisciplinares, López-Huertas (2015) destaca as peculiaridades do conhecimento interdisciplinar, a fim de contemplar estes domínios nos métodos da AD. Para abranger o conhecimento interdisciplinar na definição da AD, López-Huertas (2015, p. 574) propõe uma mudança no enunciado de Hjørland e Albertchen (1995) conforme segue:

Para estudar os domínios do conhecimento como comunidades de pensamento ou discurso, que fazem parte da divisão do trabalho da sociedade, e a nova produção de conhecimento que é expressão de um processo de integração e interação de diferentes especialidades em direção a um objeto de pesquisa em estreita relação com a sociedade.

López-Huertas (2015) aborda como a AD pode ocorrer em contextos interdisciplinares. Declara que o conceito de interdisciplinaridade ainda não está consolidado, mas traz quatro principais características:

²² No original: “Domain is an evolving and open concept that will develop as the concept is used and applied in research and practice.” (Mai, 2005, p. 605).

²³ CHRISTENSEN, Hans Dam. Kunsten at vælge fra: om bradfordsk materialevalg og vidensorganiseringens socialitet. **Dansk Biblioteksforskning**, [S. l.], v. 3, n. 1, p. 31–43, 2007. Disponível em: <https://doi.org/10.7146/danbibfor.v3i1.97548>. Acesso em: 2 maio 2024.

- a) integração e interação entre especialidades em que se reconhece um problema principal: estão intimamente ligados à comunicação e ao consenso sobre terminologia, teorias e métodos entre as especialidades sobre um objeto específico;
- b) tipo de conhecimento a ser integrado: o conhecimento científico não é o único tipo de conhecimento, sendo importante considerar a integração da sociedade na pesquisa, como forma de enriquecer o produto final;
- c) atores no processo de pesquisa: a pesquisa não é feita apenas por cientistas, podendo ocorrer fora das universidades. Esta característica se relaciona a ideia de incluir outras formas de conhecimento na interdisciplinaridade;
- d) origem do problema de pesquisa: o problema pode ser gerado de dentro da ciência (interdisciplinaridade endógena) ou pode ter origem em problemas reais da comunidade (interdisciplinaridade exógena).

Considerando o problema desta pesquisa e as características acima apresentadas, o domínio de produtos do varejo pode ser considerado interdisciplinar, à medida que se origina de um problema real, no contexto de produtos que são comercializados no varejo; envolve conhecimento não científico, com estudos e relatórios oriundos do mercado varejista; e abrange várias áreas do conhecimento, como economia, administração e marketing, e do ponto de vista do tratamento da informação, a própria CI.

López-Huertas (2015) apresenta proposta de reflexão sobre as abordagens apresentadas por Hjørland (2002), sob a perspectiva da interdisciplinaridade:

- a) indexação e recuperação: a indexação de documentos primários é uma boa forma de acessar e compilar o conhecimento, formando um corpus que representa aquela área, possibilitando a garantia literária;
- b) estudos terminológicos: a partir da indexação, é possível identificar os principais conceitos e termos na área, o que leva a terminologia daquela área. O conhecimento interdisciplinar não é tão estável quanto nas disciplinas consolidadas, mas ainda assim, é possível observar algum grupo de conceitos comuns na área, criados pela própria interdisciplinaridade, e conceitos trazidos de outras disciplinas;
- c) construção de classificações especializadas e tesouros: a análise de estruturas conceituais de uma área a partir da análise de sistemas de

classificação e tesouros revela dispersão e falta de consenso na maioria da estrutura. Esta falta de consenso é observada tanto no escopo do domínio quanto na própria estrutura do SOC. Também estas estruturas são concebidas como sistemas universais ou híbridos, com um amplo espectro temático, em vez de uma estrutura específica para o objeto em questão;

- d) estudos bibliométricos: estudos bibliométricos e cientométricos são métodos conhecidos para analisar domínios do conhecimento. Fornecem informação relevante sobre mapa temático que forma o domínio;
- e) estudos de usuários: é importante incorporar as estruturas conceituais e terminologia dos usuários aos SOCs, como por exemplo o uso de *logs* em um sistema de recuperação de informação para enriquecer o modelo;
- f) estudos documentais: analisa tipos de documentos no domínio, e os discursos neles contidos;
- g) estudos críticos e epistemológicos: a identificação de estruturas paradigmáticas já é difícil em disciplinas, quanto mais em campos interdisciplinares, que normalmente não carregam uma tradição, se originam e desenvolvem de forma diferente e não possuem um corpus de conhecimento consolidado, não compartilhando um paradigma de forma consensual.

Para a modelagem do domínio de produtos de varejo, é importante considerar a heterogeneidade de conceitos e fundamentos neste domínio, a partir de uma perspectiva interdisciplinar, visto que este domínio não se configura como uma disciplina ou uma área do conhecimento científico, mas sim, como uma especialização socialmente institucionalizada na divisão do trabalho.

3.4.2 Metodologia em Análise de Domínio

Tennis (2003) preocupa-se não em definir o que é o domínio, mas em apresentar uma abordagem metodológica sobre como definir um domínio a ser investigado, a partir de dois dispositivos analíticos, chamados de eixos, que delineiam uma definição que pode ser operacionalizada no domínio que está sendo estudado.

Primeiramente é necessário definir qual o domínio será estudado. O escopo e abrangência podem ser definidos através da especificação da extensão e da intensão do domínio que será analisado. Para especificar a extensão, deve-se nomear o domínio o mais detalhadamente possível, especificando o nível de especialização que será considerado, destacando o que ficará de fora do domínio. Portanto, o nome, a extensão e as exclusões são áreas de modulação e formam o primeiro eixo na especificação do escopo e da abrangência do domínio (Tennis, 2003):

- a) eixo 1: Áreas de modulação. Define parâmetros sobre os nomes e a extensão do domínio, que é todo seu escopo, sendo necessário nomear a extensão do domínio. Determina o que é incluído, o que não é incluído e como o domínio é chamado. No caso desta pesquisa, o domínio é nomeado como “produtos do varejo”. O produto será compreendido como um objeto com atributos que serão abstraídos em conceitos e características, que serão organizados e representados em um modelo conceitual. Portanto, fatores como a produção, comercialização e consumo destes produtos, farão parte do escopo do domínio, à medida que contribuem para a melhor compreensão dos aspectos que formam e interferem na representação destes produtos;
- b) eixo 2: Níveis de especialização. Qualifica e define a intensão do domínio. Ou seja, sua extensão é reduzida e sua intensão é aumentada. A qualificação ocorre com base em foco ou intersecção:
 - no foco, é intensificada a intensão, diminuindo sua extensão. O analista pode encontrar divisões usadas dentro do domínio para aumentar a intensão dentro da “categoria”;
 - a intersecção ocorre entre diferentes domínios, criando um novo domínio (Tennis, 2003).

Nesta pesquisa, o domínio de produtos do varejo é qualificado com foco no contexto em que foi realizada a pesquisa, ou seja, o ecossistema Hindiana, apresentado na seção 3.1 CONTEXTUALIZAÇÃO DA PESQUISA. Dentro deste contexto, é definido o recorte da pesquisa, composto por produtos do varejo alimentar, mais especificamente, biscoitos, compreendidos como produtos, do ponto de vista material e comercial. Desta forma, fica definido o domínio de conhecimento como produtos do varejo e o recorte como a categoria de biscoitos.

A finalidade da AD se refere a dois tipos de análise citados por Tennis (2012): descritivo e instrumental:

- a) a análise descritiva é utilizada por pesquisadores para descrever um domínio para fins de pesquisa básica;
- b) a análise instrumental é utilizada para modelagem de SOCs. Para isso, baseia-se em garantia literária, *feedback* dos usuários e garantia estrutural do próprio sistema.

Esta pesquisa apresenta caráter instrumental, visto que a AD será adotada como base metodológica para a estruturação de um modelo de representação, como instrumento para organização e representação de informação sobre produtos.

Ao contrário de Tennis (2003; 2012), Albrechtsen (2015) acredita que o domínio é um “campo de estudo” que é construído através do processo de planejamento, modelagem e construção de um SOC em particular, não sendo possível desenhar um planejamento prévio antes de iniciar a AD. Para exemplificar, Albrechtsen (2015) explana sobre trabalho realizado em AD, referente a reuso de *software* na década de 1990. As etapas da análise foram se desenvolvendo à medida que o trabalho ia progredindo, conforme descrito abaixo:

- a) etapa 1 – padrões e tendências de publicações: foi realizado estudo bibliométrico para analisar as publicações e assuntos tratados;
- b) etapa 2 – esboço do desenvolvimento da área e sua configuração em Ciência e Tecnologia: identificação do estado da arte sobre reuso de *software* nas principais disciplinas envolvidas (ciência da computação e engenharia de *software*);
- c) etapa 3 – análise dos interesses de conhecimento envolvidos: foram identificados paradigmas diferentes, apresentando uma vertente técnica dominante e outra vertente interessada no desenvolvimento de ferramentas para criatividade dos indivíduos e grupos;
- d) etapa 4 – análise de facetas para classificação de *software*: estudo da teoria da classificação, análise do uso de classificação facetada para *softwares* e modelagem das facetas;
- e) etapa 5 – protótipo de esquema de classificação para conceitos de *software*: construção do protótipo;
- f) etapa 6 – experimentação com o protótipo.

De acordo com Hjørland (2017), no contexto da OC, a construção de sistemas de classificação utilizando AD como abordagem metodológica, pode seguir estas etapas:

- a. Vá para um determinado domínio;
- b. Veja como é classificado de acordo com o conhecimento contemporâneo (incluindo diferentes visões);
- c. Discuta as bases, os pressupostos epistemológicos e quais os interesses atendidos pelas classificações propostas;
- d. Sugira uma classificação motivada (HJØRLAND, 2017, p. 18-19, tradução nossa).²⁴

Nesta pesquisa, as abordagens de Hjørland (2002) para AD serão adotadas da seguinte forma:

Quadro 3 – Abordagens de AD adotadas na pesquisa

Abordagem	Aplicação na pesquisa
2 - Produção e avaliação de sistemas de classificação e tesouros	Elaboração de modelo de representação de produtos do varejo, construído com base nas estruturas de SOCs.
3 - Competências em indexação e recuperação da informação	Consideração das competências em recuperação da informação dos usuários externos, varejistas de minimercados de bairro; e competências de indexação dos usuários internos, operadores que não são especializados nem no domínio do conhecimento em questão nem em indexação.
9 - Estudos terminológicos, linguagens para fins especiais e análise do discurso em áreas do conhecimento.	Perspectiva da terminologia para análise dos termos no domínio do conhecimento, incluindo variações terminológicas, visando a definição dos termos que serão adotados no modelo de representação de produtos do varejo.

Fonte: Elaborado pela autora.

A AD nesta pesquisa foi utilizada como base metodológica para o desenvolvimento de um modelo de representação, buscando uma melhor compreensão do domínio de conhecimento de produtos do varejo, por meio de revisão

²⁴ No original: “a) Go to a given domain; b) Look at how it is classified according to contemporary knowledge (including different views); c) Discuss the basis, the epistemological assumptions and which interests are served by proposed classifications; d) Suggest a motivated classification.” (Hjørland, 2017, p. 18-19).

de literatura na área, conforme apresentado na próxima seção. Com isso, pretende-se identificar os principais conceitos e como eles se relacionam no domínio de conhecimento. Este estudo serve como apoio na fase de análise de dados do corpus de pesquisa, com objetivo de validar os conceitos e relações identificadas nesta fase.

4 RESULTADOS

As próximas seções apresentam os resultados da pesquisa, iniciando pela análise do domínio de produtos do varejo, seguindo para a análise do corpus de pesquisa, demonstrando ao final o modelo de representação para produtos do varejo.

4.1 DOMÍNIO DE PRODUTOS DO VAREJO

O contexto em que foi realizada esta pesquisa é o Ecossistema Hindiana, formado por quatro empresas que atuam em diversas etapas em uma cadeia de valor, a fim de viabilizar a comercialização de produtos de grandes indústrias e distribuidores para pequenos varejistas (Hindiana, 2022). O ecossistema atua com produtos do varejo alimentar. De acordo com a Classificação Nacional de Atividades Econômicas (IBGE, 2023) esta categoria de comércio é classificada como varejo não-especializado, abrangendo comércio de mercadorias em geral, com predominância de produtos alimentícios, em hipermercados, supermercados, minimercados, mercearias e armazéns.

No ecossistema em questão, o foco são minimercados da capital paulista e região metropolitana, com perspectiva de expansão, contemplando alimentos e outros produtos de consumo básico, como produtos de limpeza, higiene, pequenos itens de bazar e ferragens. A comercialização destes produtos ocorre entre empresas, operação chamada de B2B (*Business to Business*), que se trata de um segmento do comércio eletrônico relacionado a transações *online* entre empresas, abrangendo operações de compra e venda de informações, produtos e serviços (Catalani *et al.*, 2009).

Esta transação eletrônica é realizada por meio de uma plataforma digital, em que as indústrias e distribuidores disponibilizam seus produtos para venda e os pequenos varejistas (minimercados), que possuem contrato com a empresa, compram através da plataforma digital e recebem seu pedido através das empresas de armazenamento e logística do ecossistema ou de parceiros. Esta plataforma digital pode ser compreendida como um *marketplace* que permite integrar vendedores a mercados eletrônicos, podendo ser utilizado em conjunto com outros sistemas (Vieira, 2020).

Portanto, considerando o contexto acima apresentado e o recorte de pesquisa, e buscando compreender o domínio de produtos do varejo, as próximas seções trarão referencial teórico sobre a área do varejo, produtos e biscoitos.

4.1.1 Varejo

O varejo compreende todas as atividades relacionadas à venda de bens ou serviços diretamente ao consumidor final, para uso pessoal e não empresarial. Varejista ou loja de varejo é qualquer empreendimento comercial em que o volume de vendas provenha principalmente do varejo. Qualquer organização que venda para consumidores finais está envolvida com o varejo (Kotler; Keller, 2012).

A entrega de um produto ao cliente ocorre por meio da cadeia logística, chamada também de canais de distribuição, que inicia na produção do item, com empresas produtoras, fabricantes e indústrias, passando pelo seu armazenamento, expedição, transporte, recebimento, exposição e comercialização. Os níveis de canais representam a quantidade de intermediários pelos quais um produto percorre, desde a expedição pelo fabricante até a compra pelo consumidor. Portanto, de acordo com cada produto, é necessário estabelecer uma rede de valor, chamada também de cadeia de valor, estabelecendo parcerias e alianças estratégicas capazes de contribuir para que o público-alvo perceba o valor do bem adquirido (Rocha; Mola, 2018).

Os intermediários assumem diferentes papéis na cadeia logística, sendo eles: atacadistas, distribuidores, varejistas ou atacarejos (*cash and carry*). O atacadista compra e revende aos varejistas uma grande quantidade de produtos, com preços mais baixos e com foco em produtos com grande volume de vendas, chamados de *commodities*. Os distribuidores atuam como representantes comerciais das indústrias, com foco na venda de todo o portfólio daquela indústria, direto no ponto de venda. Portanto os varejistas podem adquirir as mercadorias a partir dos atacados e dos distribuidores, e no caso de varejos maiores, podem inclusive comprar diretamente da indústria. Os varejistas representam o ponto de venda, espaço no qual o consumidor tem contato direto com o produto, tanto bens duráveis quanto bens de consumo, sendo o último elo da cadeia logística. Já o atacarejo é um modelo híbrido que agrupa as funções do atacado e do varejo em um mesmo estabelecimento comercial. Ou seja, o atacarejo vende tanto para varejista quanto para consumidores finais (Rocha; Mola, 2018; Pereira; Barros, 2017).

Os varejistas podem ser divididos pelos seguintes tipos: lojas de varejo, varejo sem loja e organizações de varejo (Kotler; Keller, 2012). Dentre as lojas de varejo, as principais são as lojas de especialidade, loja de departamentos, supermercado, loja de conveniência, farmácia, loja de desconto, loja de ultradesconto ou valor mínimo, varejista *off-price* (liquidação), superloja e *showroom* de vendas por catálogo. O varejo sem loja pode ser dividido em venda direta, marketing direto (incluindo telemarketing e vendas pela *internet*), máquinas de venda automática e serviços de compra. Já as organizações de varejo, ou varejo corporativo, possuem maior poder de compra e reconhecimento de marca do que lojas independentes. Os principais tipos de varejo corporativo são as redes corporativas, redes voluntárias, cooperativas de varejo, cooperativas de consumidores, franquias e conglomerados de comercialização (Kotler; Keller, 2012).

Em mercados como os Estados Unidos e Europa, o varejo é segmentado da seguinte forma: bens de consumo (exceção de carros e combustíveis), carros e combustíveis e serviços oferecidos ao consumidor final. No Brasil, o IBGE segmenta o setor do varejo de uma forma um pouco diferente: varejo restrito (bens de consumo, exceto automóveis e materiais de construção), varejo ampliado (todos os bens de consumo, incluindo automóveis e materiais de construção) e serviços em geral (Sociedade Brasileira de Varejo e Consumo, 2022).

Recentemente, com a pandemia de coronavírus, houve uma diferenciação entre o chamado varejo essencial (supermercados, farmácias e materiais de construção) e o varejo não-essencial (demais setores), sendo que os primeiros puderam operar sem restrições durante a pandemia, e os últimos tiveram que fechar as portas ou operar parcialmente. Este cenário acelerou a digitalização destas empresas, como uma alternativa para vendas de forma *online* (Sociedade Brasileira de Varejo e Consumo, 2022).

A Classificação Nacional de Atividades Econômicas (CNAE) divide o comércio varejista não-especializado da seguinte forma: comércio varejista de mercadorias em geral, com predominância de produtos alimentícios - hipermercados e supermercados; comércio varejista de mercadorias em geral, com predominância de produtos alimentícios - minimercados, mercearias e armazéns; e comércio varejista de mercadorias em geral, sem predominância de produtos alimentícios. De acordo o CNAE, hipermercados compreendem:

[...] as atividades dos estabelecimentos comerciais com venda predominante de produtos alimentícios variados e que também oferecem um gama variada de outras mercadorias, tais como: utensílios domésticos, produtos de limpeza e higiene pessoal, roupas, ferragens, etc. com área de venda superior a 5000 metros quadrados. (IBGE, 2023).

O supermercado é definido pelo CNAE da mesma forma que o hipermercado, com a diferença do tamanho, compreendendo uma área entre 300 e 5000 metros quadrados. Já os minimercados, mercearias e armazéns compreendem: “[...] as atividades dos estabelecimentos comerciais sem autoatendimento e com venda predominante de produtos alimentícios variados em minimercados, mercearias, armazéns, empórios, secos e molhados, com área de venda inferior a 300 metros quadrados.” (IBGE, 2023).

Os hipermercados, supermercados e minimercados, por comercializarem predominantemente produtos alimentícios, são estabelecimentos compreendidos no varejo alimentar. O supermercado é caracterizado por atuar com operações de autoserviço relativamente grandes, de baixo custo, baixa margem e alto volume, sendo projetados para atender a todas as necessidades de alimentação, higiene e limpeza doméstica (Kotler; Keller, 2012). Em lojas de autosserviço, o próprio consumidor se encarrega da tarefa da compra, pegando o produto da prateleira e levando até o caixa. O tamanho do supermercado normalmente é medido pelo total de itens disponíveis e total de *check-outs* (caixas) da loja. Normalmente o supermercado disponibiliza os produtos em cinco seções: mercearia, frios e laticínios, frutas e verduras, carnes e bazar (Rocha; Mola, 2018).

Os minimercados comercializam alimentos e artigos de primeira necessidade, com pouca variedade, e normalmente estão estabelecidos em vizinhanças de grande circulação e em pequenas instalações (Saab; Gimenez, 2000). Este formato de varejo apresenta maior penetração em diferentes classes econômicas. Um estudo da McKinsey & Company (Guissoni; Consoli; Rodrigues, 2013), a partir de dados da LatinPanel e Nielsen, mostra que os minimercados possuem em média 29% dos consumidores das classes A e B, 38% da classe média e 33% das classes D e E.

Dentre os tipos de marketing direto, que se trata de um dos tipos de varejo sem loja, está a venda *online* por meio de *e-commerce*, em que os esforços se voltam exclusivamente ao consumidor de forma direta e sem intermediários (Rocha; Mola, 2018). Em um contexto mais amplo, há o *e-business*, que compreende todo o rol de negócios realizados por meio da *web* (Cavalcante; Bräscher, 2014; Steimer; Luz,

2015). Trata-se de uma plataforma digital para gestão de uma organização de maneira eficaz, reduzindo custos e agregando valor ao cliente (Souza; Bremgartner, 2016). Já o *e-commerce*, é a abreviatura da expressão em inglês “*eletronic commerce*”, ou seja, “comércio eletrônico”. Refere-se a qualquer tentativa de negociação de produtos ou serviços de forma eletrônica, por meio da *web*. Portanto, o *e-commerce* é um dos tipos de negócios do *e-business* (Cavalcante; Bräscher, 2014).

Albertin (2020) define comércio eletrônico como: “[...] a realização de toda a cadeia de valor dos processos de negócio num ambiente eletrônico, por meio da aplicação intensa das tecnologias de comunicação e de informação, atendendo aos objetivos de negócio.” O comércio eletrônico não é visto apenas como uma inovação tecnológica, mas como uma inovação no modelo de negócio (Santarosa, 2016). É realizado por meio de uma plataforma digital que define como a marca personalizará a exibição dos produtos, o catálogo e os itens desejados pelos clientes, visando uma excelente experiência de compra, com objetivo de fidelizar o cliente. Estudos apontam principais aspectos que influenciam a experiência de compra em uma plataforma digital (Chakraborty; Lala; Warren, 2002): personalização, interatividade das atividades de compra, interatividade de atividade de relacionamento com usuários, informações, organização, privacidade e segurança, acessibilidade e entretenimento.

No modelo de loja virtual, o varejista dispõe de Centros de Distribuição (CDs) próprios, tendo controle sobre o próprio estoque dos produtos que são vendidos em sua plataforma digital. Porém, atualmente, muitas empresas estão mudando seus modelos de negócio, não se limitando a comercializar apenas os próprios produtos, mas possibilitando que terceiros, chamados de *sellers*, ofereçam seus produtos em sua plataforma, faturando uma porcentagem na revenda, utilizando de sua estrutura e marca de comércio eletrônico. Este modelo atua como uma espécie de *shopping center* virtual e é chamado de *marketplace* (Steimer; Luz, 2015).

Diferentes tipos de negociação podem ser realizados por meio do comércio eletrônico (Steimer; Luz, 2015) e elas são classificadas de acordo com a relação existente entre os participantes (Santarosa, 2016). O modelo B2C (*Business-to-Consumer*) é considerado o primeiro modelo de comércio eletrônico e continua sendo o modelo mais popularizado, tendo as lojas virtuais como principais representantes. No B2C a empresa vende diretamente para o consumidor final, utilizando a loja virtual ou *marketplace* como intermediário.

Já o modelo B2B (*Business-to-Business*) baseia-se na possibilidade das empresas negociarem entre si. Ou seja, envolve a venda de produtos de uma empresa para outra empresa (Cavalcante; Bräscher, 2014). As operações B2B podem ocorrer em qualquer ponto da cadeia produtiva, como produção, distribuição e revenda (Cheuiche *et al.*, 2019). Souza e Bremgartner (2016, p. 6) apontam as principais vantagens obtidas com o modelo B2B: “[...] eliminação dos gargalos de comunicações entre comprador e fornecedor; otimização do tempo, visto que o objetivo do uso de um sistema é promover a troca de informações por um meio eletrônico; o uso da internet como maior ferramenta, entre outras.” As transações em um modelo B2B costumam ser de altas quantias e altos volumes, portanto, é importante que seja construída uma relação de confiança e segurança entre as empresas envolvidas (Cheuiche *et al.*, 2019).

Há ainda outros modelos de negociação, conforme abaixo (Steimer; Luz, 2015; Cheuiche *et al.*, 2019):

- a) C2B (*Consumer-to-Business*): negociação eletrônica entre consumidores e empresas. Um exemplo são os *blogs* que produzem conteúdos direcionados para vendas em páginas de empresas;
- b) C2C (*Consumer-to-Consumer*): negociação direta entre os consumidores;
- c) B2G (*Business-to-Government*): relações comerciais entre empresas e governo, como projetos ou licitações;
- d) G2C (*Government-to-Consumer*): relação comercial eletrônica entre o governo e consumidores;
- e) B2E (*Business-to-Employee*): relação entre empregado e empregador.

Esta pesquisa está inserida em um contexto de uma empresa que opera em um modelo B2B, por meio de uma plataforma de comércio eletrônico no formato de *marketplace*, comercializando produtos do varejo alimentar. Portanto, é importante conhecer os principais aspectos do varejo e destes modelos de negócio, visando a análise deste domínio, com objetivo de desenvolver um modelo de representação de produtos comercializados neste contexto.

4.1.2 Produto

O produto é tudo que pode ser oferecido a um mercado para satisfazer alguma necessidade ou desejo, podendo incluir bens físicos, serviços, além de experiências, eventos, lugares, entre outros (Kotler; Keller, 2012). Pride e Ferrel (2016) definem produto como um bem, serviço ou ideia recebido em uma troca. De acordo com a Lei de Defesa ao Consumidor (Brasil, 1990), produto é definido como: “[...] qualquer bem, móvel ou imóvel, material ou imaterial.” Em uma visão mais ampla, do ponto de vista da área de Marketing, o produto faz parte do chamado mix de marketing, também conhecido como 4 Ps do marketing, sendo eles: produto, preço, lugar (em inglês, “*place*”) e promoção (Kotler; Keller, 2012). Nesta pesquisa, o produto é considerado como um bem material comercializado por meio de plataforma de comércio eletrônico. Nesta perspectiva, o produto é um objeto que carrega informações sobre si e estas informações são representadas na plataforma digital. Desta forma, o consumidor tem acesso ao produto por meio da representação de seus atributos.

O planejamento da oferta no mercado deve considerar os níveis do produto. Kotler e Keller (2012) apresentam cada nível, sendo que cada um deles vai agregando mais valor para o cliente. Os cinco níveis constituem a hierarquia de valor para o cliente:

- a) benefício central: o que realmente o cliente está comprando. Por exemplo, no caso de um hotel, o cliente está comprando descanso e pernoite;
- b) produto básico: é a manifestação do benefício central em um produto. No exemplo do hotel, seria um quarto de hotel, incluindo a cama, banheiro, toalhas, mesa, cômoda e armário;
- c) produto esperado: conjunto de atributos que o cliente normalmente espera de determinado produto. No exemplo do hotel, o hóspede espera no mínimo uma cama arrumada, toalhas e lençóis limpos, lâmpadas que funcionem e um certo grau de tranquilidade;
- d) produto ampliado: produto que excede às expectativas do cliente. No exemplo do hotel, poderia ser um frigobar, uma mensagem de boas-vindas, entre outros;
- e) produto potencial: abrange todos os possíveis incrementos e transformações que o produto poderia ser submetido no futuro. Neste nível, a empresa busca novas formas de satisfazer o cliente e diferenciar seu produto dos concorrentes.

Pride e Ferrel (2016) apresentam os níveis de produto de outra forma. A oferta total do produto é compreendida como o conjunto de três elementos independentes: o produto em si, suas características suplementares e seus benefícios simbólicos. O produto em si trata-se da sua utilidade fundamental, seu principal benefício, ou seja, para qual finalidade o produto foi desenvolvido e quais necessidades fundamentais do consumidor visa atender. As características suplementares oferecem atributos além da utilidade ou benefícios centrais, agregando valor ao produto, porém não são essenciais para o funcionamento do mesmo. Além disso, as características suplementares ajudam na diferenciação da marca. Já os benefícios simbólicos referem-se às vantagens percebidas pelo consumidor com base em sua experiência.

Segundo Kotler e Keller (2012), os produtos são classificados de acordo com suas características de durabilidade, tangibilidade e uso, podendo ser divididos em três grupos:

- a) bens não duráveis: tratam-se de bens tangíveis, normalmente consumidos rapidamente e comprados com frequência, como itens de higiene e alimentação. Em inglês, estes produtos são conhecidos pela sigla FMCG, que significa *Fast Moving Consumer Goods* (BÖSE, 2016);
- b) bens duráveis: são bens tangíveis, normalmente utilizados por um determinado período, como eletrodomésticos, ferramentas e roupas;
- c) serviços: são produtos intangíveis, inseparáveis, variáveis e perecíveis, normalmente exigindo maior controle de qualidade, credibilidade e adaptabilidade, como um corte de cabelo ou uma consulta médica.

Pride e Ferrel (2016) e Kotler e Keller (2012) classificam os produtos de acordo com as necessidades de quem está comprando. Portanto, os produtos são divididos em produtos ou bens de consumo, para atender necessidades pessoais e familiares; e produtos de negócios (ou bens industriais), para uso em operações em uma organização, revenda ou fabricação de outros produtos. Os bens de consumo são classificados de acordo com o comportamento de compra do consumidor, sendo divididos em:

- a) produtos de conveniência: itens baratos e comprados com frequência. Normalmente são ofertados em muitas lojas de varejo e não exigem muito planejamento de compra pelo consumidor. São divididos em bens básicos (comprados com regularidade), bens de impulso

- (comprados sem planejamento), bens de emergência (necessidade urgente, como um guarda-chuva durante uma tempestade);
- b) produtos de compra comparada: exigem maior esforço do consumidor comparando as lojas e marcas em relação a preços, características do produto, qualidade, serviços e garantia. São produtos que se espera que durem por um tempo longo e sejam comprados com menos frequência do que os itens de conveniência. Podem ser divididos em bens homogêneos (se assemelham em qualidade, mas se diferenciam em preço) e heterogêneos (diferem em certos atributos que podem ser mais importantes do que o preço);
 - c) produtos de especialidade: são itens de características únicas e, no geral, os compradores despendem de esforço considerável com o planejamento da compra e não aceitam substitutos. Neste caso, os compradores não fazem comparações, pois sabem exatamente o que querem;
 - d) produtos não procurados: produtos que o consumidor não conhece ou normalmente não pensa em comprar. São adquiridos para atender a um problema ou necessidade repentina, como uma emergência médica, reparos de automóveis ou serviços funerários.

Para que o consumidor identifique a identidade da marca, é importante que os produtos sejam diferenciados por meio de seus atributos (Kotler; Keller, 2012):

- a) forma: refere-se ao tamanho, formato ou estrutura física, por exemplo, as diferentes formas em que uma aspirina pode ser apresentada (posologia, cor, formato, invólucro);
- b) características: características variáveis que complementam a função básica do produto. Podem ser identificadas e selecionadas por meio de pesquisas com compradores;
- c) customização: produtos adaptados de acordo com informações coletadas sobre clientes, individualizando ofertas de produtos;
- d) qualidade de desempenho: nível no qual as características básicas do produto operam, podendo ter níveis de desempenho baixo, médio, alto ou superior;

- e) qualidade de conformidade: todas as unidades produzidas do mesmo produto devem ser idênticas e devem atender às especificações prometidas;
- f) durabilidade: indicador da vida útil esperada do produto sob condições naturais ou excepcionais. É um atributo importante para os bens duráveis;
- g) confiabilidade: indicador da probabilidade de um produto não falhar ou quebrar durante determinado período;
- h) facilidade de reparo: indicador da facilidade de se consertar um produto com mau funcionamento ou que não esteja funcionando;
- i) estilo: visual do produto e a sensação passada por ele ao consumidor. Neste aspecto entra o design do produto, como uma maneira consistente de diferenciar e posicionar os produtos de uma marca. O design é o conjunto das características relacionadas à aparência, sensação e funcionamento do produto sob a perspectiva do consumidor. Oferece benefícios funcionais e estéticos, recorrendo tanto ao lado racional quanto emocional do cliente.

Esta categorização de atributos de produtos, da perspectiva do marketing, pode trazer elementos para a análise dos dados nesta pesquisa, visto que o modelo de representação a ser desenvolvido deve refletir a forma como esta área se organiza. Neste sentido, os biscoitos podem ser compreendidos como bens de consumo não duráveis, mais especificamente como produtos de conveniência, podendo ser divididos em bens básicos ou bens de impulso.

Na perspectiva do varejo, os produtos não são compreendidos isoladamente, mas fazendo parte de grupos, formando um conjunto total de produtos. Neste sentido, os produtos podem ser organizados de acordo com a hierarquia abaixo, exemplificada com biscoitos (Kotler; Keller, 2012):

- a) família de necessidade: necessidade central que sustenta a existência de uma família de produtos. Exemplo: alimentos;
- b) família de produtos: todas as classes de produto que atendem a uma necessidade central. Exemplo: alimentos não perecíveis;
- c) classe de produtos: grupo de produtos que compõem uma família de produtos que apresentam uma certa coerência funcional, também conhecido como categoria de produto. Exemplo: biscoitos;

- d) linha de produtos: grupo de produtos que compõem uma classe de produtos que apresentam relação por desempenharem função similar, são vendidos para os mesmos grupos de clientes e comercializados pelos mesmos canais. Exemplo: biscoitos doces;
- e) tipo de produto: grupo de itens que compõem uma linha de produtos, compartilhando uma das diversas formas possíveis do produto. Exemplo: biscoito doce recheado;
- f) item (unidade específica de estoque ou variante de produto): unidade distinta dentro de uma marca ou linha de produtos, que se diferencia por tamanho, preço, aparência ou outro atributo. Exemplo: biscoito doce recheado de chocolate 100 g da Bauducco.

Um sistema de produto é um grupo de itens diferentes, porém relacionados, e que funcionam de modo compatível. Por exemplo, o sistema de produtos de um *smartphone* pode abranger carregadores, protetores de tela, fones de ouvidos, entre outros. O mix de produtos, também conhecido como sortimento de produtos, refere-se ao conjunto de todos os produtos vendidos por uma determinada empresa, sendo composto por diversas linhas de produtos. O mix de produtos possui quatro dimensões, conforme exemplo na tabela e lista abaixo (Kotler; Keller, 2012):

Figura 3 – Dimensões do mix de produtos

TABELA 12.2 Abrangência do mix de produtos e extensão das linhas de produtos da Procter & Gamble (incluindo a data de lançamento)					
Abrangência do mix de produtos	Extensão das linhas de produtos				
	Sabão em pó	Creme dental	Sabão em barra	Fraldas descartáveis	Lenços de papel
	Ivory Snow (1930)	Glem (1952)	Ivory (1879)	Pampers (1961)	Charmin (1928)
	Dreft (1933)	Crest (1955)	Camay (1926)	Luvs (1976)	Puffs (1960)
	Tide (1948)		Zest (1952)		Bounty (1965)
	Cheer (1950)		Safeguard (1963)		
	Dash (1954)		Oil of Olay (1993)		
	Bold (1965)				
	Gain (1966)				
	Era (1972)				

Fonte: Kotler e Keller (2012, p. 358).

- a) abrangência: quantidade de linhas de produtos oferecidas pela empresa. Exemplo: 5 linhas;
- b) extensão: número total de itens no mix de produtos. Exemplo: 20 itens;

- c) profundidade: quantidade de opções de cada produto na linha. Exemplo: o sabão em pó Tide pode ser apresentado em duas fragrâncias;
- d) consistência: o quão estreita é a relação entre as diversas linhas de produtos em relação ao uso final, requisitos de produção, canais de distribuição ou outro critério. Exemplo: as linhas de produtos da P&G são consistentes visto que todas são bens de consumo e passam pelos mesmos canais de distribuição. Porém são menos consistentes, considerando que desempenham diferentes funções aos consumidores.

Com base nestas quatro dimensões do sortimento de produtos, o varejista consegue traçar estratégias de expansão para seu negócio, podendo adicionar novas linhas de produtos, ampliando o mix; ou aumentar a extensão das linhas de produtos; pode também acrescentar mais opções de cada produto da linha, aumentando a profundidade da linha; ou ainda pode buscar maior consistência entre as linhas de produtos (Kotler; Keller, 2012).

Uma maneira de fazer a gestão do sortimento de produtos no varejo é o gerenciamento por categorias (ou gerenciamento de categorias). Trata-se de um processo em que as categorias de produtos são gerenciadas como unidades de negócios estratégicas, produzindo resultados aprimorados, com foco na entrega de valor para o cliente (Glynn, 2007; Larentis; Slongo; Milan, 2006). “Essas categorias são grupos gerenciáveis distintos de produtos que os consumidores veem como inter-relacionados e substituíveis.” (Glynn, 2007, p. 56). Pereira e Barros (2017, p. 2) definem gerenciamento de categorias como: “[...] uma ferramenta de gestão que contribui para proporcionar uma melhor exposição, um melhor abastecimento, maior variedade de marcas e tamanhos, visando reduzir os estoques e aumentar o giro do mercado.”

A categoria é composta por um grupo de produtos com finalidades e utilidades semelhantes e que atendem de maneira conjunta uma demanda específica do consumidor final (Pereira; Barros, 2017). A estrutura do conjunto dos produtos em um varejo normalmente é organizada nos seguintes cinco níveis hierárquicos: departamento, categoria, subcategoria, segmento e subsegmento (Pereira; Barros, 2017).

Esta estratégia se contrapõe ao gerenciamento por marcas, que agrupa e gerencia os produtos de acordo com sua marca e fabricante. No gerenciamento por categorias, o varejista gerencia as marcas como parte de uma categoria de produtos, otimizando o lucro geral da categoria (Glynn, 2007). No Brasil, esta estratégia é amplamente utilizada em super e hipermercados, porém os minimercados apresentam dificuldades para adesão a esta ferramenta, devido à falta de recursos e pessoal adequados para implantação, ou até mesmo devido à falta de interesse ou conhecimento por parte dos varejistas (Pereira; Barros, 2017).

O gerenciamento por categorias atua como uma ferramenta de colaboração e relacionamento entre fabricantes e varejistas (Glynn, 2007). Porém, cabe ressaltar que a decisão final de implantação do gerenciamento por categorias é do varejista. Cabe ao fornecedor apresentar informações de mercado e as categorias de forma clara e transparente, para que as decisões sejam tomadas com segurança e objetividade (ECR Brasil, 2007). Alguns dos benefícios do gerenciamento por categorias para os varejistas são o melhor desempenho da categoria e gerenciamento de estoque (Glynn, 2007). Os principais benefícios para os fabricantes são a integração das estratégias dos clientes da marca e do varejista, podendo se concentrar à categoria de forma mais estratégica com os varejistas, e não apenas enfatizar margem de lucro e preço (Glynn, 2007).

O gerenciamento por categorias nasceu no ECR (*Efficient Consumer Response*), programa surgido nos Estados Unidos, em 1992, visando maior eficiência das operações entre fabricantes e varejistas, otimizando seus processos e aumentando a satisfação dos clientes. A partir dos anos 2000, na perspectiva do gerenciamento de categorias, o cliente é tratado como *shopper*. O *shopper*, diferentemente do consumidor, é a pessoa que realiza a compra na loja, independentemente de quem vai consumir o produto adquirido (Pereira; Barros, 2017). As estratégias do gerenciamento por categorias devem estar alinhadas com as estratégias da empresa, com foco no *shopper* (Pereira; Barros, 2017).

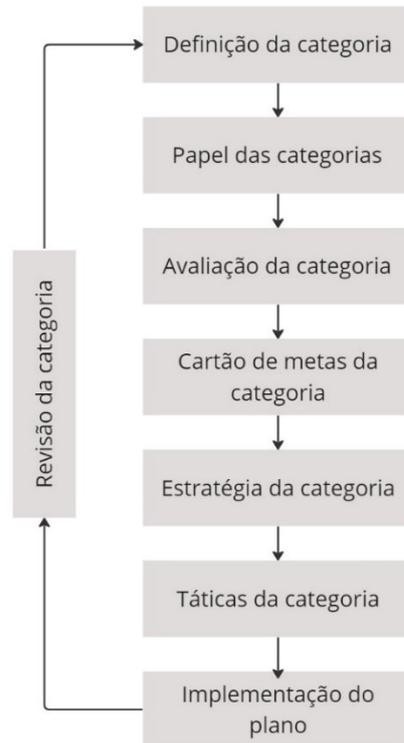
Com toda a evolução da internet e das lojas que estão cada vez menores e mais completas, o *shopper* vem ficando cada vez mais informado e exigente no momento da compra. Com isso, o varejista tem de ir adaptando suas ofertas para o cliente, assim poderá utilizar ferramentas para que tenha um diferencial dentro do mercado, tentando concorrer com os grandes supermercados e até com sites da internet. (Pereira; Barros, 2017, p. 7).

No ECR, quatro ações são fundamentais: reposição eficiente de produtos, sortimento eficiente de produtos, promoção eficiente de produtos e introdução eficiente de novos produtos (Larentis; Slongo; Milan, 2006). Dentre as técnicas que compõem o ECR, o gerenciamento de categorias se apresenta como uma alternativa para gerenciar o sortimento de produtos de maneira mais eficiente. O processo do gerenciamento de categorias é dividido em etapas, conforme abaixo (Larentis; Slongo; Milan, 2006; Pereira; Barros, 2017):

- a) definição da categoria: determinação dos produtos que fazem parte da categoria, subcategoria e segmentos, com base nas necessidades e comportamento de compra do *shopper*;
- b) papel da categoria: atribuição de papel de cada categoria, com base na comparação multicategoria, informações sobre consumidores, mercado e concorrentes;
- c) avaliação da categoria: análise das subcategorias e os segmentos que compõem a categoria, com base em informações sobre consumidores, mercado, fabricantes e concorrentes;
- d) cartão de metas da categoria: estabelecimento dos indicadores de desempenho e metas de cada categoria, com base em sua avaliação;
- e) estratégias da categoria: realização do papel da categoria e consecução dos objetivos definidos no cartão de metas;
- f) táticas da categoria: consideração do mix de marketing (4 Ps) de cada categoria, de forma que assegure a execução da estratégia e o alcance dos objetivos e metas, utilizando árvore de decisão do *shopper*;
- g) implementação do plano: definição de cronograma e determinação de responsabilidades;
- h) revisão da categoria: realização de avaliações periódicas permanentes e monitoramento.

Estas etapas podem ser visualizadas na figura abaixo:

Figura 4 – Processo de gerenciamento por categorias



Fonte: Elaborado pela autora.

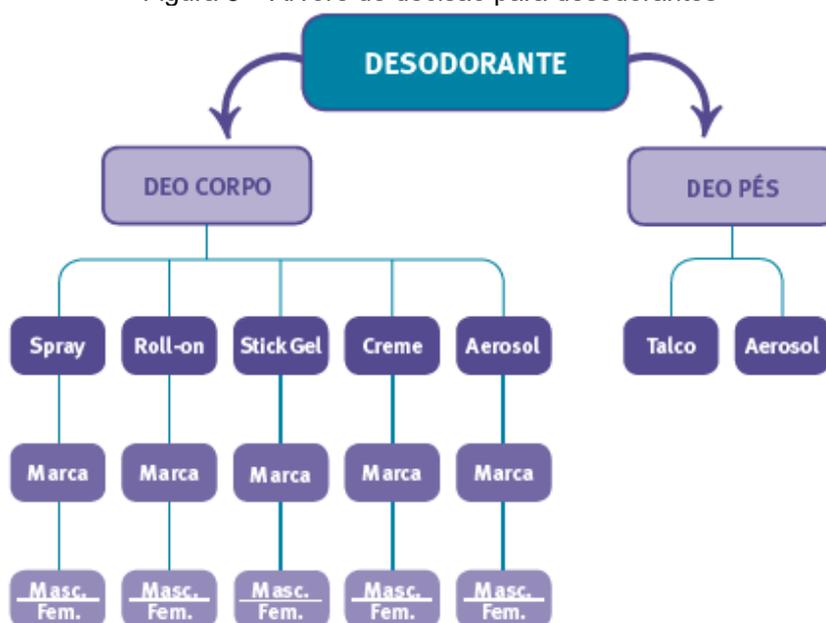
Existem quatro papéis de categoria, conforme abaixo (ECR Brasil, 2007). Cada varejista oferece uma combinação destes papéis de acordo com sua proposta, perfil demográfico e estilo de vida de seu público-alvo:

- a) destino: papel pelo qual a loja é lembrada como melhor local para comprar determinado produto. Não significa apresentar o melhor preço, mas sim, o melhor valor percebido pelo *shopper*, composto por sortimento, preço, ambientação e serviço;
- b) rotina: papel que define o varejista como loja preferida para fornecimento de valor consistente e competitivo, para atendimento de necessidades rotineiras do *shopper*;
- c) ocasional/sazonal: papel que reforça imagem do varejista como loja preferida para fornecimento de valor adequado ao consumidor, no atendimento de demandas sazonais;
- d) conveniência: papel que reforça a imagem do varejista como loja completa para fornecimento de valor adequado, no atendimento de necessidades de reposição planejadas e não-planejadas.

Para a etapa de definição da categoria, é necessário conhecer o processo decisório do *shopper*. O resultado deste processo configura-se na árvore de decisão

de compra (ECR Brasil, 2007). A partir do conhecimento dos hábitos do *shopper*, é possível identificar os critérios pelos quais ele escolhe um determinado produto entre os demais, permitindo a construção da árvore de decisão da categoria. Na figura abaixo, é apresentado um exemplo de árvore de decisão para a categoria “desodorante” (ECR Brasil, 2007, p. 30):

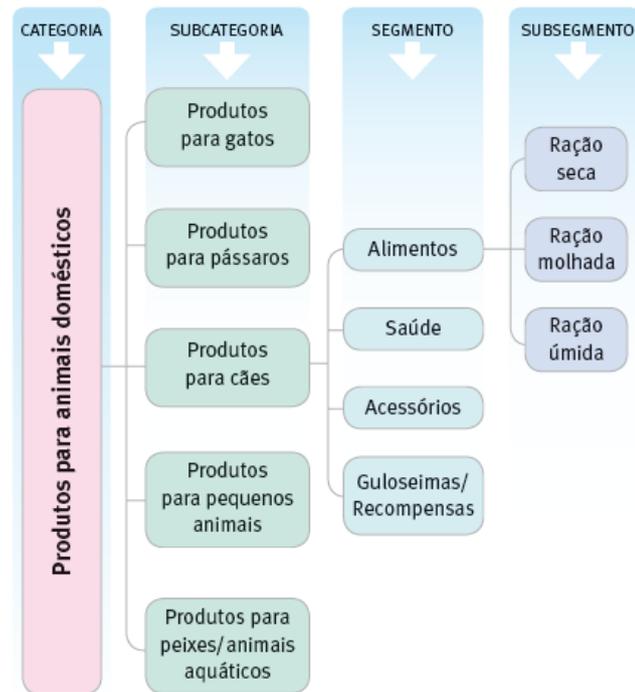
Figura 5 – Árvore de decisão para desodorantes



Fonte: ECR Brasil (2007, p. 30).

Neste exemplo, é possível perceber que o principal critério que determina a escolha do *shopper* é o aplicador, se é spray, roll-on etc. Após, a marca é considerada na escolha e, por último, se o desodorante é feminino ou masculino. Portanto, evidencia-se que a árvore de decisão é uma representação dos critérios de escolha do *shopper* (ECR Brasil, 2007) e é ferramenta importante para a estruturação das categorias de produtos. Com base na árvore de decisão, é possível criar uma estrutura mercadológica conforme figura abaixo (ECR Brasil, 2007, p. 13):

Figura 6 – Exemplo de estrutura mercadológica



Fonte: ECR Brasil (2007, p. 13).

A estrutura mercadológica é a “[...] a ferramenta do varejo que define estrategicamente o sortimento de cada categoria.” (ECR Brasil, 2007, p. 31). Ela é uma estrutura viva, não estática, estando sujeita a contínuas alterações de acordo com revisões e movimentações do mercado (ECR Brasil, 2007). Assim, como visto anteriormente, uma estrutura hierárquica, como uma taxonomia, é composta por categorias e subcategorias, formando cadeias e renques. Cada categoria na cadeia é subdividida em subcategorias mais específicas, com base em um critério de subdivisão. É possível observar a semelhança entre a taxonomia e a estrutura mercadológica. Para esta última, é possível utilizar os critérios de escolha do *shopper* identificados na árvore de decisão, para subdividir as categorias de produtos na estrutura mercadológica, visto que, ela deve ser organizada para atender as necessidades do *shopper*.

ECR Brasil (2007) destaca a importância da qualidade do cadastro de produtos para análise de desempenho das categorias. É necessário que cada produto contenha a sua respectiva classificação de acordo com a estrutura mercadológica. Além disso, deve-se evitar problemas resultantes de descrições diferentes para o mesmo produto. Por exemplo, se um mesmo amaciante for primeiramente cadastrado como “amac” e depois como “amaciante”, isto gerará dois registros diferentes no cadastro para o mesmo produto, podendo impactar análises de negócio e emissão de documentos

fiscais. Por este motivo, ECR Brasil (2007) ressalta a importância da padronização da descrição de produtos utilizando pelo menos quatro variáveis: produto (exemplo: amaciante de roupas), marca (exemplo: Comfort), detalhe (exemplo: natural) e conteúdo (exemplo: 500 ml).

No ambiente digital, o principal fator de sucesso do gerenciamento de categorias é a fidelidade e retenção dos clientes, sendo capazes de impulsionar indicadores-chaves de desempenho para o varejo como número de categorias na cesta, valor médio da cesta e frequência média de visitas à loja virtual. No comércio eletrônico, apesar dos compradores terem maior facilidade de alternar entre lojas, é mais provável que eles permaneçam fiéis a uma loja em comparação com o varejo físico. Para isso, a experiência geral de compra deve ser consistente e positiva para o usuário. Os varejistas de bens de consumo rápido, como alimentos, bebidas, artigos de higiene e limpeza, ainda se mostram cautelosos a participar plenamente do comércio eletrônico, pois se preocupam com a canibalização de seus canais físicos. Porém, estudos mostram que os canais de venda *online* e *offline* costumam se complementar (Böse, 2016).

Böse (2016) apresenta estudo sobre as diferenças do gerenciamento de categorias em lojas virtuais e lojas físicas, compreendendo que o comércio eletrônico não é apenas mais um canal de vendas, mas sim um novo método de distribuição com especificidades operacionais diferentes das lojas físicas. A maneira como o consumidor compra é diferente na loja virtual, portanto as táticas de gerenciamento de categorias devem ser adaptadas. No ambiente digital, não há corredores pelos quais o *shopper* percorre e pode ser atraído por alguma marca ou embalagem. No digital, o item único torna-se mais relevante, assemelhando-se ao modelo de catálogo. O contato do *shopper* com o produto é limitado à dimensão “cognitivo-afetiva”, ou seja, a interação com o produto é delimitada pelas informações que são fornecidas sobre ele (Böse, 2016), que, pela perspectiva da OC, trata-se da representação da informação. Portanto, torna-se ainda mais importante a facilidade de acesso e clareza das informações disponíveis sobre os produtos ofertados (Santarosa, 2016).

A definição da categoria na loja virtual pode ser mais flexível, pois os produtos costumam ser alocados de acordo com a ocasião de uso, podendo o mesmo produto ser alocado simultaneamente em várias posições diferentes no menu da loja virtual (Böse, 2016). Da perspectiva da OC, poderiam ser consideradas algumas alternativas para dar conta desta necessidade. Uma das possibilidades em uma taxonomia ou

tesauro é a polihierarquia, para que o mesmo produto possa ser classificado em categorias de cadeias diferentes (por exemplo, um panetone poderia ser classificado tanto em Mercearia quanto em Sazonais). Porém, esta alternativa não é recomendada no contexto desta pesquisa, visto que a polihierarquia necessita maior esforço e conhecimento técnico para sua adequada manutenção. Portanto, poderia se analisar a utilização de taxonomias facetadas, ou ainda, classificar hierarquicamente em apenas uma categoria, mas estabelecer uma relação associativa com outra categoria relevante. Esta última opção poderia ser utilizada na adoção de tesauros ou ontologias.

Quanto à recuperação da informação, Böse (2016) destaca que o comprador que busca por um item específico, normalmente utiliza a barra de pesquisa na página da loja virtual, pesquisando por meio de palavras-chaves. Já o consumidor que não tem uma definição do produto que está procurando, costuma navegar pelas opções do menu de navegação. O resultado da pesquisa depende do mecanismo de busca adotado, porém é importante que seja apresentado por ordem de relevância para o consumidor (Böse, 2016). Na perspectiva da OC, para a melhoria da busca, é importante a utilização de remissivas, para que, mesmo que o usuário não pesquise pelo termo exato adotado pela loja, ele consiga recuperar o produto da mesma forma. Já na navegação, pode ser utilizada uma taxonomia com estrutura das categorias dos produtos. Os atributos dos produtos, formalizados por meio de metadados no sistema, podem ser apresentados no resultado da busca como forma de refinamento da pesquisa, atuando como filtros (por exemplo, na categoria de camisetas femininas, poderia haver filtros para cor, tamanho, material etc.). Böse (2016) destaca que a conveniência e a facilidade de pesquisa contribuem imensamente para a experiência e fidelidade do cliente.

No varejo digital, como o consumidor não tem contato direto com o produto, o processo de devolução e substituição se torna mais relevante. Em relação à substituição, caso não haja estoque ou o produto procurado não esteja disponível, é importante apresentar alternativas, de acordo com a árvore de decisão da categoria, com base no critério de decisão menos relevante ao consumidor (Böse, 2016). Na perspectiva da OC, poderiam ser apresentadas alternativas de produtos classificados em categorias coordenadas ou na categoria imediatamente superior. Também poderia se avaliar o uso de relações associativas para atender a esta necessidade, além de

considerar atributos utilizados para a representação descritiva do produto (por exemplo: outras camisetas com as mesmas características, mas de outra cor).

Outro aspecto que interfere na experiência e na decisão de compra do cliente na loja virtual, são as avaliações de outros consumidores (Böse, 2016). Em relação a este aspecto, pela perspectiva da OC, poderiam ser utilizados os metadados do produto como parâmetros para que os usuários relatem a sua avaliação, fornecendo, desta forma, uma informação mais precisa e relevante para os demais usuários que irão consultá-la (por exemplo, se na avaliação consta que a camiseta recebida na compra é menor do que o anunciado, é importante para quem irá consultar esta avaliação, saber qual o tamanho da camiseta comprada por quem avaliou).

Em relação a promoção de produtos, Böse (2016) destaca que uma forma de impulsionar as vendas é utilizar as recomendações personalizadas de produtos diretamente na área de *check-out*. Em relação a este aspecto, pela perspectiva da OC, poderiam ser trabalhadas as relações coordenadas, hierárquicas e associativas da categoria do produto na cesta de compras. Porém, neste caso, também é importante a utilização de algoritmos de recomendação baseados nas experiências anteriores dos usuários, ou seja, a possibilidade de também apresentar recomendações com base no histórico de compras anteriores.

Em síntese, o conjunto de produtos comercializados por um varejista corresponde ao seu sortimento de produtos. A gestão deste sortimento pode ser realizada por meio do gerenciamento de categorias. No contexto do comércio eletrônico, o sortimento da loja virtual é disponibilizado aos consumidores por meio das informações sobre os produtos, que são representadas pela sua categoria e atributos. O conjunto das representações dos produtos do sortimento é chamado de catálogo de produtos. Da perspectiva de gestão de produto, para o gerenciamento das informações sobre os produtos neste catálogo, há uma área chamada *Product Information Management* (PIM).

O PIM se refere aos processos e tecnologias destinados ao gerenciamento centralizado de informações sobre produtos, com foco nos dados necessários para comercializar e vender os produtos por meio de um ou mais canais de distribuição (Abraham, 2014). O PIM ganhou espaço com o crescimento do comércio eletrônico, que trouxe a necessidade de o vendedor coletar dados básicos e claros sobre os produtos. Além disso, em lojas virtuais, é possível ofertar uma quantidade maior de produtos em comparação com as lojas físicas, devido à flexibilidade de estoque. Com

isso, houve um crescimento na quantidade de produtos vendidos no comércio eletrônico, agravando a necessidade de gerenciamento deste grande volume de dados. Os dados dos produtos podem ser acessados por diversos canais, como os dispositivos móveis, adicionando mais uma camada de complexidade para a gestão destes dados. Os consumidores, por sua vez, demandam por informações mais completas e precisas sobre os produtos. Na compra *online*, o cliente busca e compara por diversos detalhes dos produtos, antes de efetuar a compra. Além disso, o cliente espera que haja consistência da informação entre os diversos canais disponíveis (Abraham, 2014).

No PIM, os dados de produtos são centralizados e armazenados em um único local, chamado de “*single source of truth*”, que distribui a informação para sistemas de informação. O processo do PIM é composto pelas seguintes etapas (Abraham, 2014):

- a) coleta: importação, mapeamento e transformação dos dados;
- b) consolidação: limpeza, mesclagem e consolidação dos dados;
- c) enriquecimento: enriquecimento dos dados por meio de categorização, classificação e adição de mídias;
- d) distribuição: mapeamento, customização e exportação.

Na etapa de enriquecimento dos dados, são acrescentadas mais informações sobre o produto ao seu cadastro, além das informações básicas coletadas e consolidadas nas etapas anteriores do processo. Dentre as informações acrescentadas, estão as informações relativas à categorização e classificação. A categorização, na perspectiva do PIM, permite que a organização classifique os produtos de um catálogo em grupos específicos, propiciando uma visão geral melhor de todo sortimento de produtos. A estrutura das categorias é conhecida como “árvore de categorias”, “árvore de produtos” ou “taxonomia de produtos”.

Observa-se a relação entre o conceito de árvore de categorias com o conceito de estrutura mercadológica empregado no gerenciamento por categorias. Ambos se referem a uma estrutura composta por categorias de produtos, organizadas hierarquicamente, ou seja, uma taxonomia. Porém, no caso do gerenciamento de categorias, ela é utilizada com foco no desempenho comercial da categoria e no PIM é utilizada como forma de organizar as informações sobre produtos, com foco no enriquecimento dos dados do catálogo de produtos.

O sistema PIM pode permitir que o produto seja classificado apenas no último nível da taxonomia ou em níveis superiores (Abraham, 2014). Esta é uma decisão de

negócio importante, visto que muitas vezes as descrições recebidas para o cadastro de produtos são insuficientes para classificá-lo no nível mais específico da estrutura. Além disso, o sistema pode permitir que o produto seja classificado em mais de uma categoria ou apenas em uma (Abraham, 2014). Esta definição pode facilitar o acesso do usuário ao produto por meio de outras categorias que podem fazer sentido na busca em determinado contexto. Porém, sem uma estruturação adequada, pode resultar em classificação cruzada. Esta definição pode impactar também a análise sobre o desempenho das categorias e sortimento, visto que a contabilização de produtos por categoria resultará em um valor maior do que a quantidade total de produtos no sortimento.

Já a classificação se refere ao gerenciamento de atributos. Também conhecida como taxonomia ou esquema. Os atributos descrevem um elemento específico de um determinado produto, como nome, tamanho, cor etc. No PIM, são determinados quais os atributos serão utilizados como padrão para a representação de determinado tipo de produto (Abraham, 2014). Na perspectiva da OC, a definição dos atributos por tipo de produto está relacionada com a identificação de características e definição dos metadados para representação descritiva deste. A adoção de um padrão de metadados por tipo de produto somado a uma estrutura de categorias de produtos (como uma taxonomia, tesouro ou ontologia) possibilita com que os metadados sejam definidos de acordo com cada categoria nos níveis da árvore, fazendo com que as categorias subordinadas recebam como herança os atributos das categorias superordenadas, tornando mais consistente o uso dos atributos e facilitando a gestão dos mesmos.

Uma das funcionalidades do PIM é a criação de relacionamentos entre produtos. Abraham (2014) cita algumas finalidades para estes relacionamentos, como: mostrar produtos similares, oferecer um produto de maior valor, oferecer acessórios ao produto pesquisado, explicar que um item faz parte de um conjunto de produtos, oferecer peças de reposição ou informar sobre obrigatoriedade de compra de um item junto a outro item, substituir um produto antigo por um mais novo etc. Na perspectiva da OC, estes relacionamentos podem se dar por meio de relações associativas entre conceitos. Os sistemas de PIM também permitem que sejam adicionados ao cadastro do produto, fotos ou outros tipos de mídia, como vídeos e manuais do produto (Abraham, 2014).

Em um sistema PIM, as diferentes apresentações físicas de um mesmo produto são chamadas de variantes (conhecidas também como itens ou artigos). As variantes existem fisicamente e possuem um preço e nível de estoque. Um mesmo produto, como um refrigerante Coca-Cola, pode ser apresentado em garrafa PET em 1,5 litros ou 2 litros, por exemplo. Neste caso, para o sistema PIM, é criada uma hierarquia em que o produto é o refrigerante Coca-Cola que apresenta variações em dois volumes. No nível de produto da hierarquia, são atribuídos metadados como ingredientes e descrição do produto. Estes metadados são herdados pelas variantes, que recebem outros metadados referentes a sua unidade, como ID e volume (Abraham, 2014).

Atualmente há estudos e iniciativas sobre a representação de produtos e o desenvolvimento de catálogos de produtos no ambiente digital. Chen, Li e Zhang (2010) propõem o desenvolvimento de catálogos eletrônicos integrados e inteligentes, que possibilitam consulta semântica na *web* e recomendação inteligente de produtos e serviços em mercados eletrônicos B2B e C2C, com base em padrões internacionais de classificação de produtos e protocolos da *web* semântica, incluindo ontologias. Guo e Sun (2003) apresentam uma abordagem colaborativa de representação de produtos chamada COPE (*Collaborative Product Representation for Emergent Electronic Marketplace*). Essa abordagem permite que os usuários capturem mudanças em tempo real e melhorem a eficiência da plataforma digital B2B. O COPE é implementado por meio de uma linguagem de modelagem de ontologia. Matos e Belfo (2022) apresentam proposta de solução centrada em ontologia para gerenciar informações de produtos para sistemas modulares complexos disponíveis no catálogo de uma organização multinacional do setor de engenharia e tecnologia. As taxonomias existentes no catálogo de produtos *online* em diferentes sistemas foram analisadas e foi criada uma ontologia utilizando a ferramenta Protégé e validada por especialistas.

A informação sobre os produtos é um direito do consumidor e dever das empresas, especialmente no comércio eletrônico, onde não há o contato direto do cliente com os produtos, apenas com as representações dos mesmos, por meio das plataformas digitais (Steimer; Luz, 2015). É possível observar que existe no mercado varejista a compreensão da importância da informação e da organização de produtos para a melhoria do desempenho de vendas e da gestão no varejo. Há processos, ferramentas e estratégias na área da gestão do varejo e de produtos para lidar com isso, e que perpassam por conceitos da OC, como taxonomia, atributos de produtos, catálogos, entre outros. Portanto, fica evidente a possibilidade de contribuição de um

estudo interdisciplinar entre estas áreas, visando a melhoria do processo de organização e representação da informação de produtos do varejo, bem como a possibilidade de ampliação da atuação profissional e da pesquisa para a área de OC.

4.1.3 Biscoitos

O biscoito, também conhecido como bolacha, consiste em um produto obtido pela mistura de farinhas, amidos ou féculas com outros ingredientes, submetidos a processos de amassamento e cocção, fermentados ou não, podendo apresentar cobertura, recheio, formato e textura diversos (Agência Nacional de Vigilância Sanitária, 2022). O Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (2008) destaca que outras substâncias alimentícias podem ser incorporadas aos biscoitos, como a gordura hidrogenada. A partir desta definição, é possível perceber que o biscoito abrange fatores como ingredientes, processos de produção, além de aspectos como sabor, cobertura, recheio, formato e textura. Do ponto de vista do varejo, o biscoito é tratado como um produto alimentício, que é definido pela Agência Nacional de Vigilância Sanitária (2000) como: “[...] todo alimento derivado de matéria-prima alimentar ou de alimento "in natura", adicionado, ou não, de outras substâncias permitidas, obtido por processo tecnológico adequado”.

No Brasil, os biscoitos estão presentes na alimentação de mais de 90% dos lares no país. Os biscoitos doces recheados, biscoito água e sal e *cream cracker*, biscoitos secos doces e especiais, biscoito maria e maisena, rosquinhas, biscoito *wafer*, biscoitos salgados e *cookies* representam cerca de 99% das vendas em toneladas, sendo que os biscoitos doces recheados e os biscoitos *cream crackers*, juntos, representam a metade do mercado nacional (Associação Brasileira das Indústrias de Biscoitos, Massas Alimentícias e Pães & Bolos Industrializados, 2020; Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008). O mercado brasileiro de biscoitos vem sofrendo transformações importantes desde a década de 1990. A abertura para importação do trigo incentivou as indústrias nacionais a investirem neste mercado. Por outro lado, a abertura do mercado ocasionou o aumento da concorrência internacional, especialmente com a importação de biscoitos argentinos. Empresas que até o momento não participavam deste mercado, como indústrias de massas e moinhos, começaram a produzir biscoitos, resultando em um

crescimento de 50% entre 1994 e 1997 (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

O aumento da concorrência e o enfraquecimento da economia após 1997, fez com que as empresas buscassem maior ênfase em diferenciação, por meio de lançamentos de novos produtos, com novas formulações e diversificação de ingredientes. Além disso, foram lançadas novas marcas, tamanhos, formatos e embalagens e novas funções nutricionais para os biscoitos. O mercado se tornou extremamente diversificado e competitivo, contando hoje com cerca de 200 tipos diferentes de biscoitos, em um mercado altamente especializado (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

O aumento da oferta junto à retração do mercado entre 1997 e 2001, obrigou as empresas a reduzirem suas margens, diminuindo o preço médio dos produtos. Como não houve redução do preço dos principais ingredientes dos biscoitos (farinha de trigo e gordura vegetal), as empresas buscaram ganhos com o aumento da produtividade. Além disso, vem ocorrendo um efeito chamado “inflação disfarçada”, provocado pela pressão de empresas multinacionais líderes no mercado de biscoitos. Estas empresas conseguiram ganho no preço médio por meio da redução disfarçada do peso das embalagens dos produtos, sem uma redução proporcional do preço. Ou seja, desde 2000 as empresas estão reduzindo o conteúdo das embalagens de biscoitos, sem reduzir o preço. Esta prática das lideranças no mercado fez com que as empresas menores fossem forçadas a seguir o mesmo caminho (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

Outro movimento do mercado é a oferta de produtos de maior valor e sofisticação direcionados a segmentos que demandam linhas diferenciadas. Foram lançados novos biscoitos com cobertura de chocolate, tortinhas com recheio aparente, variações de sabores com apelo regional, entre outros, que aceitaram bem as variações de embalagens, especialmente as embalagens individuais. Estas últimas, passaram a representar 3/4 das embalagens fabricadas, sendo preferidas até mesmo para consumo familiar. Isto se deve a razões como: satisfação de necessidades individuais de cada membro da família, estímulo a experimentação de novidades e menor desembolso unitário (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

O mercado brasileiro de biscoitos mostra uma concentração dos oito maiores fabricantes somando 49% da participação no mercado, sendo estes: M. Dias Branco,

Nestlé, Marilan, Nabisco, Bauducco, Bom Gosto, Mabel e Piraquê (Ferreira, 2007). Os grandes fabricantes têm buscado espalhar suas fábricas por diversas regiões do Brasil, com objetivo de reduzir custos de distribuição e garantir que os produtos cheguem a locais mais longínquos. A região nordeste tem sido alvo de interesse neste sentido, especialmente por ser o maior polo consumidor de biscoitos no país (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

O Brasil é um dos maiores produtores de biscoitos no mundo, apresentando oportunidades de expansão em vendas no setor. Algumas alternativas a serem exploradas são a diversificação das linhas, tipos, sabores e embalagens dos produtos, além da identificação de nichos de mercados não explorados pelos concorrentes (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

A cadeia produtiva do biscoito compreende um longo canal que liga desde as matérias-primas utilizadas na produção até o produto final, para os consumidores finais. Esta cadeia é dividida em 4 etapas (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008):

- a) matérias-primas: envolve o cultivo e produção de matérias-primas, como farinha, leite, ovos etc., além de matérias-primas químicas utilizadas em conjunto para elaboração dos biscoitos (estabilizantes, conservantes etc.);
- b) produção: envolve os fornecedores de máquinas e equipamentos, indústria de embalagens, capacitação de pessoal, universidades e centros de tecnologia com pesquisas e desenvolvimento de novas formulações, instituições financeiras com oferta de crédito, estruturas institucionais e regras sanitárias que regulamentam e fiscalizam o setor;
- c) comercialização/distribuição: envolve padaria e produtores domésticos que comercializam os produtos diretamente em seus pontos de venda, atacados e varejos, restaurantes e bares, exportadores, empresas de logística.
- d) consumidor final: local, regional, nacional ou internacional.

O biscoito é produzido basicamente com quatro ingredientes: manteiga ou margarina, farinha de trigo, açúcar e ovos. De acordo com Agência Nacional de Vigilância Sanitária (2000), ingrediente é: “[...] qualquer substância, incluídos os aditivos alimentares, empregada na fabricação ou preparação de um alimento e que

permanece no produto final, ainda que de forma modificada.” Os ingredientes utilizados na fabricação de biscoitos podem ser divididos em duas funções principais: amaciadores (açúcar, gordura, emulsificante e fermento biológico) e estruturadores (farinha, sal, leite, água e ovos). Podem ser produzidos biscoitos macios e úmidos ou biscoitos crocantes e secos, em formato mais fino e delicado (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

Os biscoitos podem ser formados e cortados por diversos processos, de acordo com seu tipo. A massa é misturada e depois passa para a fase de formação do biscoito. Os biscoitos estampados (maria, maisena, sortidos) normalmente são laminados, passando por entre dois rolos, atingindo o estampador à medida que vai afinando a massa. Já em relação ao corte, pode ser dividido em três tipos, sendo eles: sistema de corte por prensa (estampadores), utilizados para massas duras; sistema rotativo (corte por rolos); e sistema de fios (corte por fio-aramé) (Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008).

Os biscoitos podem ser classificados de diversas maneiras, considerando critérios como, por exemplo, teor de proteína, processo de fabricação, processo de fermentação etc. O Serviço Brasileiro de Respostas Técnicas (2007) define três tipos de biscoitos:

- a) biscoitos de massa dura: conhecidos como estampados, são produtos com baixo teor de proteína, como os biscoitos maria e maisena;
- b) biscoitos de massa mole: com teor de proteína médio, como rosquinhas;
- c) biscoitos de massa fermentada: com teor de proteína elevado, como os biscoitos *cream cracker* e água e sal.

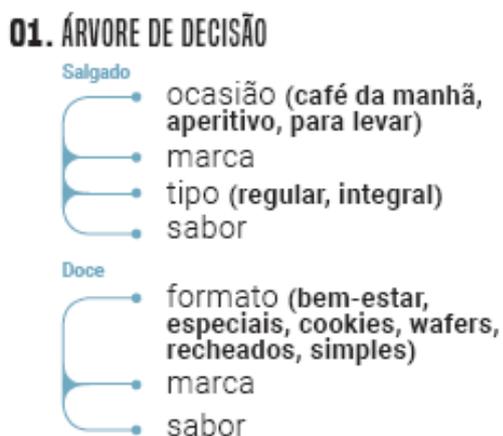
Monteiro e Martins²⁵ (2003 *apud* Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008) classificam os biscoitos de acordo com o processo de fabricação: doces recheados, semi doces (maria e maisena), doces (moldados e estampados), *crackers* (*creams cracker* e água e sal), *wafers*, *snacks* e extrusados. Segundo o Sindicato da Indústria de Massas Alimentícias e Biscoitos no Estado de São Paulo (Sindicato da Indústria de Massas Alimentícias e Biscoitos no Estado de

²⁵ MONTEIRO, A. R. G.; MARTINS, M. F. Processo de desenvolvimento de produtos na indústria de biscoitos: estudos de casos em fabricantes de médio porte. *In*: CONGRESSO BRASILEIRO DE GESTÃO DE DESENVOLVIMENTO DE PRODUTOS, 4., 2003, Gramado. **Anais** [...]. Porto Alegre: FEENG, 2003.

São Paulo ²⁶, 2007 *apud* Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2008), os biscoitos podem ser classificados da seguinte forma: biscoitos salgados, doces, recheados, revestidos, grissini, biscoitos para aperitivos e petiscos ou salgadinhos, palitos para salgadinhos ou *pretzel*, *wafer*, *wafer* recheado e *petit four*.

O SA Varejo (2019) apresenta os dados de pesquisa sobre desempenho de categorias, divididos nos seguintes grupos: mercearia de alto giro, mercearia doce, matinais, perecíveis congelados, perecíveis lácteos, bebidas não alcoólicas, bebidas alcoólicas, higiene e beleza, limpeza, descartáveis e bazar. A categoria de biscoitos é encontrada no grupo de mercearia doce. O SA Varejo (2019, p. 34) apresenta árvore de decisão para biscoitos, com base em dados do fabricante Mondelez, conforme figura abaixo:

Figura 7 – Árvore de decisão para biscoitos



Fonte: SA Varejo (2019, p. 34).

O NCM (2021), Nomenclatura Comum do Mercosul, é um sistema ordenado que permite determinar um único código numérico para uma dada mercadoria. Este código é utilizado em todas as operações de comércio exterior dos países do Mercosul. No NCM, o biscoito é classificado na seguinte cadeia, do mais geral ao mais específico: Preparações à base de cereais, farinhas, amidos, féculas ou leite; produtos de pastelaria > Produtos de padaria, pastelaria ou da indústria de bolachas e biscoitos, mesmo adicionados de cacau; hóstias, cápsulas vazias para medicamentos, obreias, pastas secas de farinha, amido ou fécula, em folhas, e produtos semelhantes > Bolachas e biscoitos, adicionados de edulcorantes; *waffles* e *wafers*. Subordinadas a

²⁶ SINDICATO DA INDÚSTRIA DE MASSAS ALIMENTÍCIAS E BISCOITOS NO ESTADO DE SÃO PAULO. Legislação. Site institucional. São Paulo, 2007. Disponível em: <http://www.simabesp.org.br/legis.asp#alim>. Acesso em: 12 abr. 2008.

esta última categoria estão: Bolachas e biscoitos, adicionados de edulcorante; e Waffles e *wafers*.

Na próxima seção serão apresentados SOCs e padrões de metadados já existentes no domínio de produtos do varejo e de biscoitos, enquanto produtos alimentícios, que poderão servir de referência para o desenvolvimento do modelo de representação de produtos do varejo nesta pesquisa.

4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento

Nesta seção, serão apresentados os principais padrões de metadados e SOCs encontrados na literatura, direcionando interesse especial aos sistemas para representação de produtos e de biscoitos. De maneira geral, esses instrumentos consistem em uma estrutura taxonômica, que pode incluir ou não atributos vinculados a classes e categorias, e podem ser formalizados por meio de metadados.

4.1.4.1 Schema.org

O Schema.org é um vocabulário controlado genérico e extensível, desenvolvido em conjunto por companhias de mecanismos de busca, dentre elas o Google, Microsoft (Bing), Yahoo e Yandex, sendo lançado em 2010. Sua manutenção, promoção e atualização é realizada por uma comunidade colaborativa, através de plataformas de fórum na *web*. Seu objetivo é padronizar e simplificar a estruturação de dados na *web*, por meio de um vocabulário padronizado e compartilhado entre mecanismos de busca, otimizando a recuperação da informação (Schema.org, 2023d; Machado; Arakaki, 2022).

A grande quantidade e diversidade de linguagens adotadas para estruturar dados em páginas na *web*, faz com que os desenvolvedores tenham que lidar com diferentes codificações, que podem levar a termos incorretos e duplicações. Portanto, a iniciativa do Schema.org propõe a adoção de seu vocabulário, com objetivo de possibilitar maior facilidade na estruturação dos dados e, desta forma, atender os requisitos da *web* semântica, podendo ser uma ponte para dados conectados (Schema.org, 2023d; Machado; Arakaki, 2022).

Conforme Barker e Campbell (2014) se os metadados do Schema.org estiverem disponíveis, qualquer mecanismo de pesquisa será capaz de

processá-los para otimizar sua interface de pesquisa, distinguindo entre recursos diferentes com o mesmo nome; permitindo que as informações mais relevantes sejam exibidas com destaque na página de resultados; possibilitando que os resultados sejam filtrados por propriedades como, preço, fornecedor ou data de publicação, além de propiciar links para localização de resultados relacionados ao mesmo assunto ou editor. (Machado; Arakaki, 2022, p. 15-16).

O Schema.org é formado por esquemas de metadados compostos por termos genéricos, padronizados e organizados em uma estrutura hierárquica formada por classes, propriedades e enumerações. “O vocabulário atualmente consiste em 803 tipos, 1465 propriedades, 14 tipos de dados, 87 enumerações e 463 membros de enumeração.” (Schema.org, 2023b, tradução nossa).²⁷

De acordo com Schema.org (2023a, tradução nossa), a estrutura do vocabulário está organizada da seguinte maneira:

1. Temos um conjunto de tipos, organizados em uma hierarquia de herança múltipla, onde cada tipo pode ser uma subclasse de vários tipos.
2. Temos um conjunto de propriedades:
 - a) cada propriedade pode ter um ou mais tipos como seus domínios. A propriedade pode ser usada para instâncias de qualquer um desses tipos.
 - b) cada propriedade pode ter um ou mais tipos como seus intervalos. Os valores da propriedade devem ser instâncias de pelo menos um desses tipos.²⁸

A estrutura do vocabulário se inicia com a classe *Thing*, mais genérica, sendo subdividida nas seguintes subclasses: *Action*, *BioChemEntity*, *CreativeWork*, *Event*, *Intangible*, *MedicalEntity*, *Organization*, *Person*, *Place*, *Product* e *Taxon*. A classe “produto”, mais diretamente relacionada com esta pesquisa, é definida pelo Schema.org como: “Qualquer tipo de produto ou serviço oferecido” (Schema.org, 2023c, tradução nossa).²⁹ As classes subordinadas à classe *Thing* possuem suas próprias subclasses e propriedades. Estas propriedades podem receber dados dos seguintes tipos: valor, enumeração ou entidade. Estes dados são organizados em outra hierarquia, e são compreendidos como *DataTypes* e abrangem os seguintes tipos de dados: *Boolean*, *Data*, *DateTime*, *Number*, *Text* e *Time* (Machado; Arakaki, 2022).

²⁷ No original: “The vocabulary currently consists of 803 Types, 1465 Properties 14 Datatypes, 87 Enumerations and 463 Enumeration members.” (Schema.org, 2023b).

²⁸ No original: “1. We have a set of types, arranged in a multiple inheritance hierarchy where each type may be a sub-class of multiple types. 2. We have a set of properties: a) each property may have one or more types as its domains. The property may be used for instances of any of these types. b) each property may have one or more types as its ranges. The value(s) of the property should be instances of at least one of these types.” (Schema.org, 2023a).

²⁹ No original: “Any offered product or service.” (Schema.org, 2023c).

A estrutura do Schema.org foi projetada para representar entidades mais relevantes para os fundadores e que otimizassem a interpretação pelos mecanismos de buscas (Machado; Arakaki, 2022). Apesar da estrutura ser bastante genérica, é possível desenvolver extensões a partir do *core* do Schema.org, devido a sua flexibilidade, funcionando como camadas sobre a estrutura principal. Desta forma, é possível desenvolver uma extensão externa, mas com base no Schema.org, para representar objetos digitais em um domínio do conhecimento especializado ou para determinados tipos de recursos informacionais. “Desse modo, o vocabulário possibilita a reutilização e adaptação dos metadados, conforme as necessidades de contextos informacionais distintos.” (Machado; Arakaki, 2022, p. 9).

O Schema.org é compatível com diferentes codificações, como *Resource Description Framework in Attributes* (RDFa), *JavaScript Object Notation for Linked Data* (JSON-LD) e Microdados (MACHADO; ARAKAKI, 2022). De acordo com a página do Schema.Org, mais de 10 milhões de sites utilizam o vocabulário para estruturar suas páginas na *web* e mensagens de *e-mail*, além de aplicativos como Google, Microsoft, Pinterest e Yandex que utilizam o vocabulário em suas experiências (SCHEMA.ORG, 2023d). Em mecanismos de busca, como o Google, os dados estruturados são enriquecidos semanticamente e são integrados a tecnologias como *Rich Snippets* e *Knowledge Graphs*, auxiliando o mecanismo de busca na coleta, interpretação e classificação das informações. Os *Rich Snippets* são elementos criados automaticamente a partir de dados estruturados constantes em páginas na *web*, integrando principais informações sobre um determinado recurso ou entidade, gerando uma visualização enriquecida de conteúdo no próprio resultado da busca (Machado; Arakaki, 2022).

O Google fornece diretrizes para a utilização do Schema.org em páginas *web*, destacando quais as classes e propriedades são obrigatórias para que o mecanismo de busca recupere dados de uma página com resultado enriquecido (Machado; Arakaki, 2022). O Google também fornece diretrizes específicas para utilização do Schema.org, particularmente da classe “produto” com objetivo de enriquecer os dados sobre o produto e melhorar sua posição e visualização nos resultados de busca (Kejriwal *et al.* 2021). Para o caso de produtos, o Google (2024b) indica como propriedades obrigatórias *name* (nome do produto), *image* (URL da foto do produto) e *offers* (oferta de venda do produto).

Para Kejriwal *et al.* (2021) o Schema.org apresenta um padrão de metadados importante para plataformas de comércio eletrônico. O projeto *Web Data Commons* (WDC) tem rastreado páginas na *web* que utilizam o padrão de metadados Schema.org, inclusive no domínio de produtos e comércio eletrônico (KEJRIWAL *et al.* 2021).

O Schema.org se mostra como um vocabulário controlado formado por esquemas de metadados importante para a estruturação de dados em páginas na *web*, especialmente para objetos e conteúdos comerciais, sendo compatível com os principais mecanismos de busca no mercado.

4.1.4.2 GS1

O GS1 é uma associação sem fins lucrativos, que tem como missão emitir códigos de barras exclusivos para produtos. O foco do GS1 são dados de produtos para comercialização e logística. Os produtores fornecem os dados ao GS1, decidindo quem poderá ter acesso. Cerca de 1.400 empresas fornecem dados ao GS1 (Top; Timmer; Simsek-Senel, 2020).

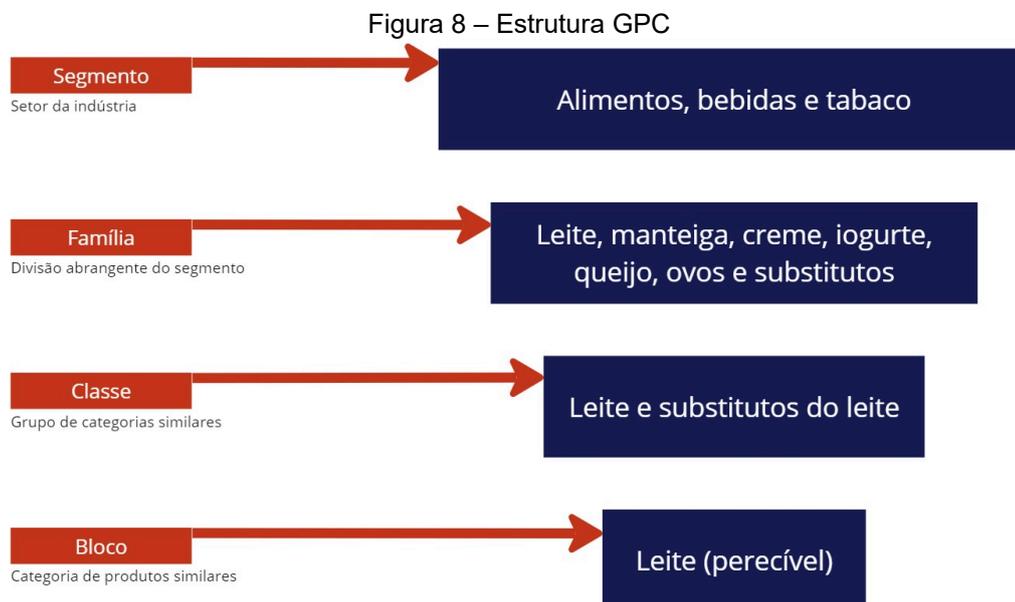
A entidade disponibiliza a Rede Global de Sincronização de Dados (GDSN), maior rede de dados de produtos do mundo:

Com o GDSN, o conteúdo de produtos de alta qualidade é carregado, mantido e compartilhado automaticamente, garantindo que os parceiros comerciais tenham acesso imediato às informações mais atuais e completas necessárias para trocar produtos nos mercados locais e globais. (TOP; TIMMER; SIMSEK-SENEL, 2020, p. 6, tradução nossa).³⁰

O GS1 desenvolve e participa de iniciativas para a organização destas informações, como o UNSPSC (*United Nations Standard Products and Services Code*), que oferece um sistema único de classificação global de produtos e serviços, em cinco níveis, gerenciado pelo GS1 norte-americano. Neste sistema de classificação, os biscoitos são subordinados à categoria de “Pães e produtos de padaria”, apresentando as seguintes subcategorias: “Biscoitos saborosos simples” e “Biscoitos ou bolinhos doces” (2022). Esta classificação não é muito precisa ou específica. Em contrapartida, a classificação GPC (*Global Product Classification*) do

³⁰ No original: “With GDSN, high quality product content is uploaded, maintained and shared automatically, ensuring trading partners have immediate access to the most current and complete information needed to exchange products on both local and global markets. (Top; Timmer; Simsek-Senel, 2020, p. 6).

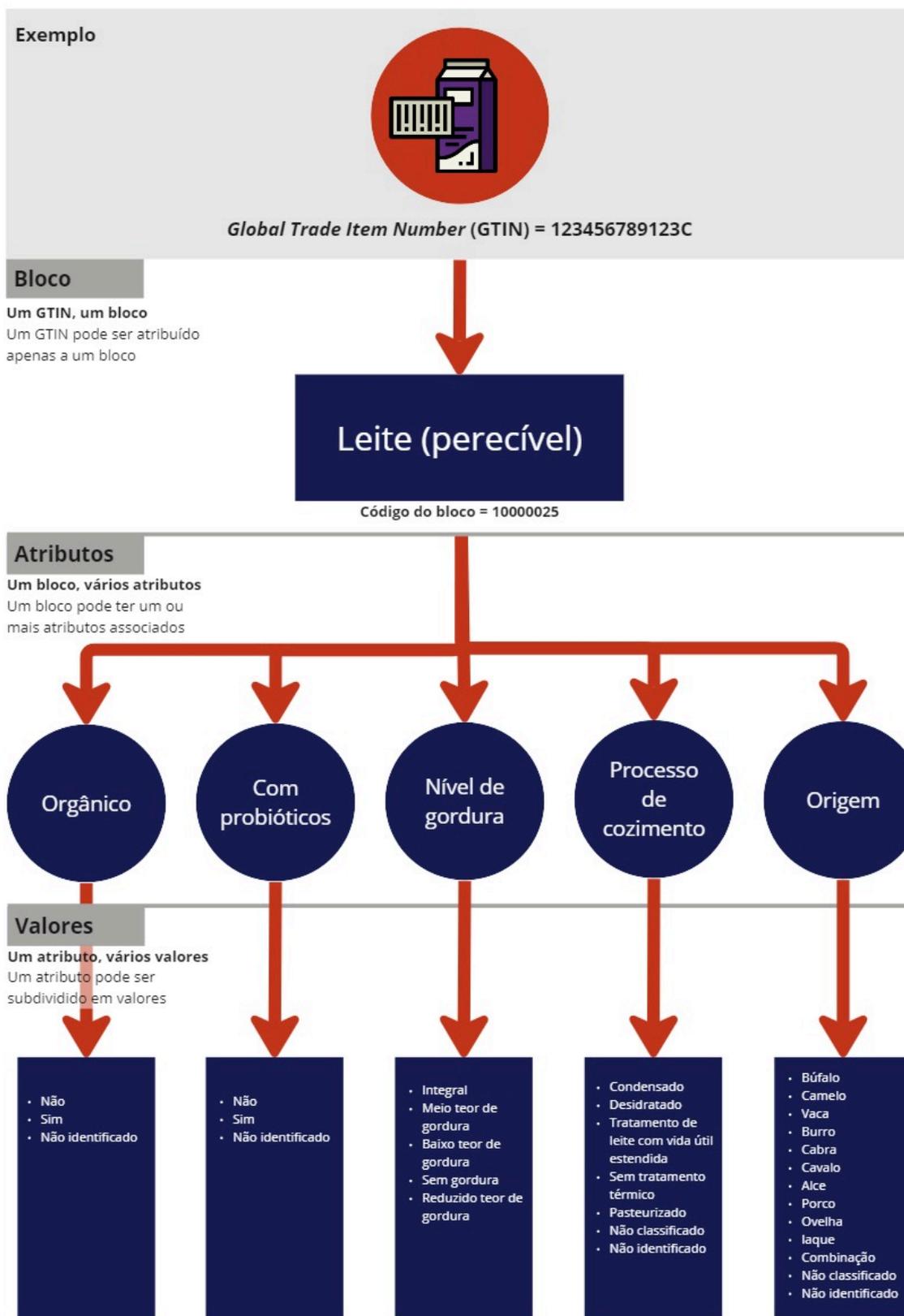
GS1 apresenta uma estrutura hierárquica mais detalhada para a organização das informações sobre produtos, dividida em quatro níveis, sendo eles: segmento, família, classe e bloco, conforme mostra figura abaixo:



Fonte: Adaptado de GS1 (2024a)

Para cada bloco (*brick*), o GPC apresenta os atributos aplicáveis àquele bloco e seus valores disponíveis, conforme mostrado no exemplo abaixo:

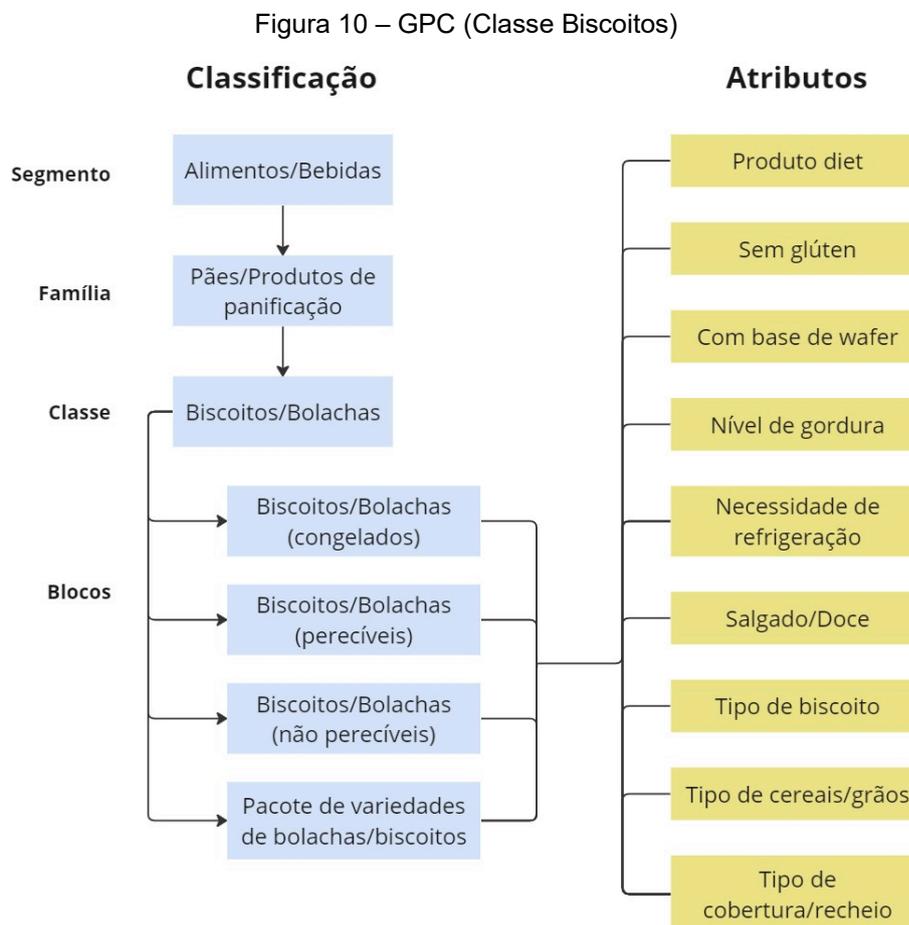
Figura 9 – Atributos GPC



Fonte: Adaptado de GS1 (2024a).

No GPC, as classes e subclasses relacionadas a biscoitos são mais específicas do que as classes apresentadas no UNSPSC, além de apresentar também seus

respectivos atributos e valores de atributos (GS1, 2024a). A figura abaixo ilustra a estrutura do esquema de classificação e lista os atributos relacionados aos blocos da classe de biscoitos:



Fonte: Elaborado pela autora.

O esquema de classificação do GPC é utilizado no *GS1 Web Vocabulary*, que reúne termos definidos em vários padrões e sistemas de dados do GS1, disponibilizados para uso geral, seguindo os princípios de *Linked Data*. O *GS1 Web vocabulary* foi projetado como uma extensão externa do padrão de metadados Schema.org, abrangendo os conceitos de Produto, Oferta e Organização, estendendo-os com propriedades mais detalhadas. O principal foco do vocabulário são as propriedades voltadas para consumidores de roupas, calçados, alimentos, bebidas e tabaco, além de propriedades comuns a todos os produtos (GS1, 2024b).

De acordo com Schema Blog (2016), o vocabulário do GS1 se baseia em um extenso conjunto de padrões *Business-to-Business* (B2B) pré-existentes. Mesmo que possa haver alguma divergência entre as terminologias do GS1 e do Schema.org,

é compartilhada uma abordagem comum fundamentada no vocabulário principal do Schema.org e nos padrões fundamentais do W3C, como a codificação JSON-LD.

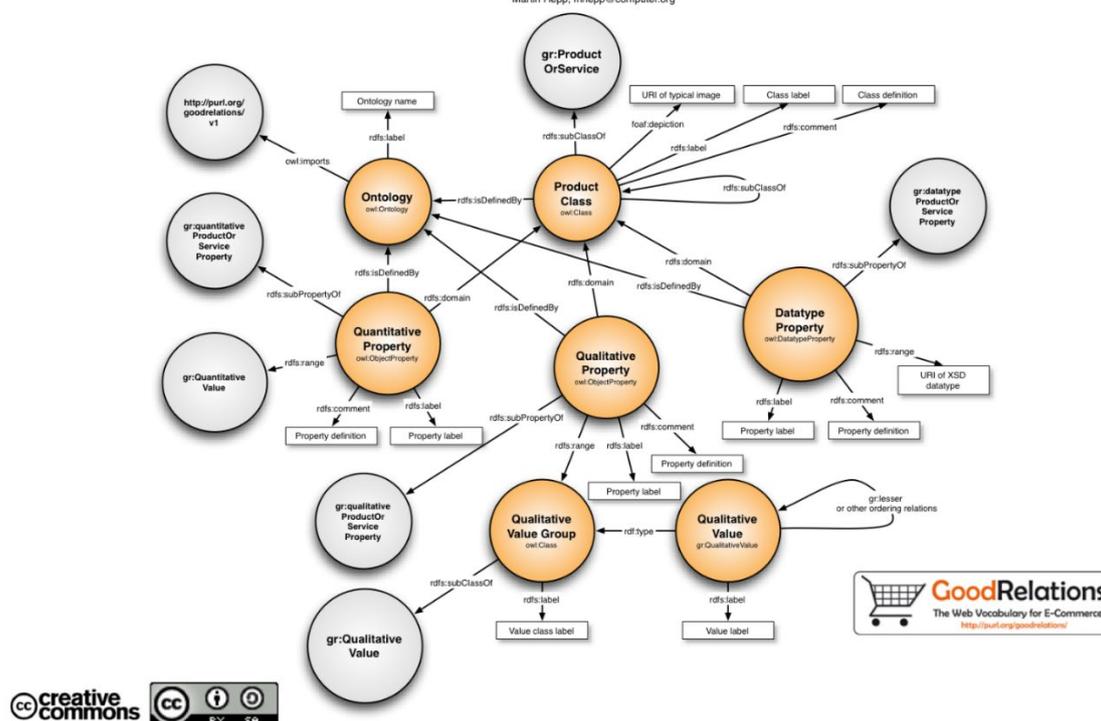
Comparando a classe Produto do Schema.org com a mesma classe do GS1 *Web Vocabulary*, identifica-se que o primeiro apresenta 66 propriedades, enquanto o segundo lista 94 propriedades representadas como metadados. Isto revela como o GS1 *Web Vocabulary* apresenta mais possibilidades para a descrição da classe Produto, em comparação ao Schema.org. Além disso, a classe de Produto no GS1 *Web Vocabulary* apresenta subdivisão para produtos alimentícios, bebidas e tabaco, com propriedades ainda mais específicas que podem ser utilizadas para descrição de biscoitos.

Por se tratar de uma instituição destinada ao registro de produtos, que coleta e disponibiliza dados sobre produtos para empresas, é importante considerar a utilização do GPC e do GS1 *Web Vocabulary* para a representação de produtos em plataformas de comércio eletrônico. Desta forma, é possível utilizar um vocabulário em conformidade com o padrão de metadados Schema.org, que é adequado aos mecanismos de buscas na web, e também ter a disposição metadados mais especializados do que no Schema.org, pois o GPC e o GS1 *Web Vocabulary* detalham as classes do Schema.org referentes a produtos.

4.1.4.3 GoodRelations

GoodRelations é um vocabulário composto por uma estrutura de dados genérica, destinado a descrição de produtos e serviços para comércios eletrônicos. O vocabulário está disponível desde 2008 e apresenta uma cobertura conceitual de todos os elementos essenciais de informações para comércio eletrônico. Ele é formado por 27 classes (tipos de entidade), 49 propriedades de objetos, 43 propriedades de dados e 43 valores enumerados (indivíduos) (Ashraf *et al.*, 2011; Goodrelations Wiki, 2023b). O vocabulário GoodRelations está disponível como uma ontologia, de acordo com o padrão W3C OWL, podendo ser utilizado para projetos de *web* semântica. O vocabulário pode ser codificado em diferentes sintaxes como Microdados, RDF, JSON-LD, XML, entre outros (Goodrelations Wiki, 2023b). GoodRelations apresenta axiomas de subclasses e subpropriedades para expressar o comportamento de subsunção no modelo (Ashraf *et al.*, 2011). A figura abaixo apresenta a estrutura do GoodRelations (2023a):

Figura 11 – Estrutura do GoodRelations
Pattern for GoodRelations-compliant Product Ontologies



Fonte: GoodRelations (2023a).

Por ser uma estrutura de dados genérica, o GoodRelations pode ser utilizado por qualquer tipo de indústria, em qualquer posição na cadeia de valor, em qualquer país. É adequado para intercâmbio de informações entre entidades comerciais (B2B) e entre empresas e usuários finais (B2C). GoodRelations é uma extensão do Schema.org para comércio eletrônico que pode ser utilizada na descrição de produtos em páginas de comércio eletrônico para aumentar a visibilidade das informações para mecanismos de busca, além de apresentar efeitos positivos para página no resultado da pesquisa, como o Google *Rich Snippets*. A adoção do vocabulário GoodRelations também traz benefícios para a gestão de informações sobre produtos dentro de uma organização ou em uma cadeia de valor, podendo servir como um esquema de banco de dados global para registrar as informações sobre produtos. GoodRelations também pode ser utilizado para gerenciamento de qualidade de dados de comércio eletrônico, gerenciando dados técnicos ou comerciais de fontes heterogêneas em bancos de dados de grafos do conhecimento (Goodrelations Wiki, 2023b).

No GoodRelations, a classe de produto e serviço é definida como:

A superclasse de todas as classes que descrevem tipos de produtos ou serviços, por natureza ou finalidade. Exemplos dessas subclasses são "aparelho de TV", "aspirador de pó", etc. Uma instância dessa classe pode

ser um produto ou serviço real (gr:Individual), uma instância de espaço reservado para instâncias desconhecidas de uma mercadoria produzida em massa (gr:SomeItems), ou uma especificação de modelo/protótipo (gr:ProductOrServiceModel). (Goodrelations Wiki, 2023c).

A classe “Produtos e Serviços” é subdividida em três subclasses: indivíduo, modelo de produto ou serviço, ou alguns itens. Estas subclasses não apresentam nenhuma subclasse subordinada a elas. Por ser um modelo genérico, ele não especifica tipos de produto em suas classes, porém, o GoodRelations indica a possibilidade de construção de extensão de outros vocabulários com base em sua estrutura, apresentando diretrizes básicas para seu desenvolvimento (Goodrelations Wiki, 2023a).

4.1.4.4 Product Type Ontology

A *Product Type Ontology* é uma ontologia OWL desenvolvida com apoio do governo alemão, pelo grupo de pesquisa E-Business e Web Science, da Universidade de Bundeswehr Munich. A *Product Type Ontology* representa o domínio de produtos e serviços, fornecendo cerca de 300 mil definições para tipos de produtos ou serviços. A ontologia estende os padrões de metadados Schema.org e GoodRelations para representação de informações sobre produtos em páginas de comércio eletrônico (Hepp, 2023b).

De acordo com Vajda, Merrell e Smith (2019), a *Product Types Ontology* foi projetada para que cresça de forma análoga a Wikipedia. Ou seja, os termos são acrescentados à ontologia diretamente das páginas da Wikipedia, com base nas solicitações dos usuários. Este modelo é motivado pela ideia de que os termos incluídos são relevantes para os usuários e se referem a entidades reais. Porém, devido a esta estratégia, o vocabulário resultante não é controlado por especialistas em ontologia ou do domínio em questão. Isto torna a estrutura confusa e fraca do ponto de vista semântico formal.

Uma das classes mais genéricas da ontologia é “Produto ou Serviço”, que, para Vajda, Merrell e Smith (2019), pode se tornar problemático para um alinhamento com ontologias de alto nível, pois nestas ontologias a distinção entre continuantes (entidades que permanecem no tempo) e ocorrentes (entidades que ocorrem no tempo) são fundamentais. Este é um aspecto que se observa também em algumas

das ontologias e padrões de metadados apresentados nesta seção, em que a classe mais genérica para produto costuma ser representada como produto e/ou serviço.

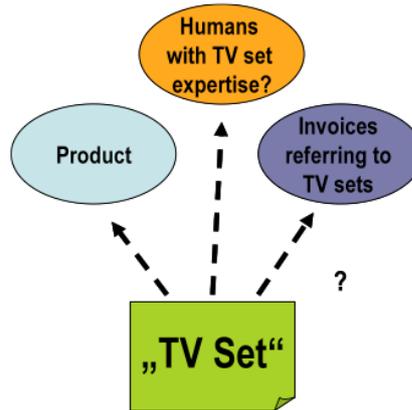
Vajda, Merrell e Smith (2019) também destacam que as propriedades constantes na *Product Types Ontology*, como qualidades, funções, datas, entre outras, não apresentam definições sistemáticas para seus termos. Um aspecto positivo são os sinônimos em vários idiomas, visto que a ontologia busca ser amplamente utilizada na *web*.

4.1.4.5 eClassOWL

A eClassOWL é uma ontologia OWL criada para descrever os tipos e propriedades de produtos e serviços na *web* semântica. Esta ontologia é totalmente compatível com o padrão de metadados Schema.org, e deve ser utilizada de forma combinada com o mesmo, pois o Schema.org contempla os aspectos comerciais de ofertas e demanda, por exemplo, preços, pagamento ou opções de entrega, diferentemente da eClassOWL que representa os tipos de produtos. A eClassOWL, assim como a *Product Types Ontology*, é um projeto desenvolvido pelo grupo de pesquisa E-Business e Web Science, da Universidade de Bundeswehr Munich (Hepp, 2023a).

A eClassOWL é estruturada em duas classes. Uma delas é denominada “genérica” e representa produtos ou serviços reais, sendo subdividida por tipos de produtos. A outra classe é denominada “taxonômica” e representa um conjunto mais amplo de objetos que podem estar relacionados a essa categoria. Além disso, as mesmas categorias podem ser usadas para descrever diferentes tipos de objetos. De forma objetiva, ao se descrever um objeto, utiliza-se o conceito genérico para produtos reais ou modelos de produtos e o conceito taxonômico para qualquer coisa que não seja um produto nem um modelo de produto (Hepp, 2023a). Este aspecto pode ser observado na figura abaixo:

Figura 12 – Representação de produtos e assuntos na eClassOWL



Fonte: Hepp (2023a).

A eClassOWL suporta três tipos de propriedades de produto: propriedades quantitativas, para características com um intervalo numérico; propriedades qualitativas, para recursos com instâncias de valor pré-definido; e propriedades de tipos de dados, como texto, data ou booleano (Hepp, 2023a). A classe de produtos e serviços é subdividida em três subclasses, sendo estas, indivíduo, modelo de produto ou serviço, ou alguns itens, da mesma forma em que ocorre no padrão de metadados GoodRelations.

4.1.4.6 LanguaL

LanguaL é um tesouro multilíngue destinado à descrição de dados sobre alimentos, especificamente para classificar produtos alimentícios, para recuperação de informações. O desenvolvimento do LanguaL teve início nos Estados Unidos no final dos anos 1970, sendo empregado tanto nos Estados Unidos quanto na Europa para manter bancos de dados numéricos sobre a composição de alimentos, incluindo informações sobre nutrientes e contaminantes. Desde 1996, a gestão do tesouro é realizada pelo *European LanguaL Technical Committee*. Atualmente, há mais de 40.000 termos sobre alimentos da Europa, da América do Norte e de outros países indexados no LanguaL (Top; Timmer; Simsek-Senel, 2020).

O tesouro é baseado nos seguintes princípios (LanguaL, 2023):

- a) todo alimento ou produto alimentício pode ser descrito sistematicamente por uma combinação de características (facetadas);
- b) essas características podem ser categorizadas em diferentes pontos de vista e codificadas para serem processadas computacionalmente;

c) estes códigos podem ser usados para recuperar dados sobre alimentos de bancos de dados externos.

O LanguaL é estruturado como uma classificação facetada em que cada alimento é descrito por um conjunto de termos controlados escolhidos a partir de características (facetas) nutricionais ou sobre a qualidade higiênica de um alimento, como por exemplo a origem biológica, os métodos de cozimento e conservação, o formato físico, as embalagens e os tratamentos tecnológicos. Antes mesmo da criação do RDF/SKOS, o LanguaL foi estruturado de forma independente, utilizando identificadores abstratos baseados em códigos únicos. Cada código único é associado a termos equivalentes em diversos idiomas (Top; Timmer; Simsek-Senel, 2020).

O tesauro não fornece atributos específicos para categoria de produtos, apenas a estrutura taxonômica de cada faceta. Os biscoitos estão categorizados na faceta “Tipo de produto”, conforme figura abaixo:

Figura 13 – Estrutura taxonômica da categoria de biscoitos



Fonte: LanguaL (2023).

O EFSA FOODEX2 é um sistema de classificação e descrição de alimentos que foi incluído no LanguaL em sua versão de 2017 (Top; Timmer; Simsek-Senel, 2020). Observa-se que o LanguaL oferece uma classificação mais detalhada para categoria de biscoitos. Assim como o GPC, do GS1, o LanguaL classifica biscoitos em uma categoria mais abrangente de produtos de panificação ou padaria, indicando a classe de produtos a qual os biscoitos pertencem, da perspectiva do próprio domínio de conhecimento. Porém, ao contrário do GPC, o LanguaL subdivide a categoria de biscoitos em biscoitos doces ou planos e biscoitos recheados ou com cobertura,

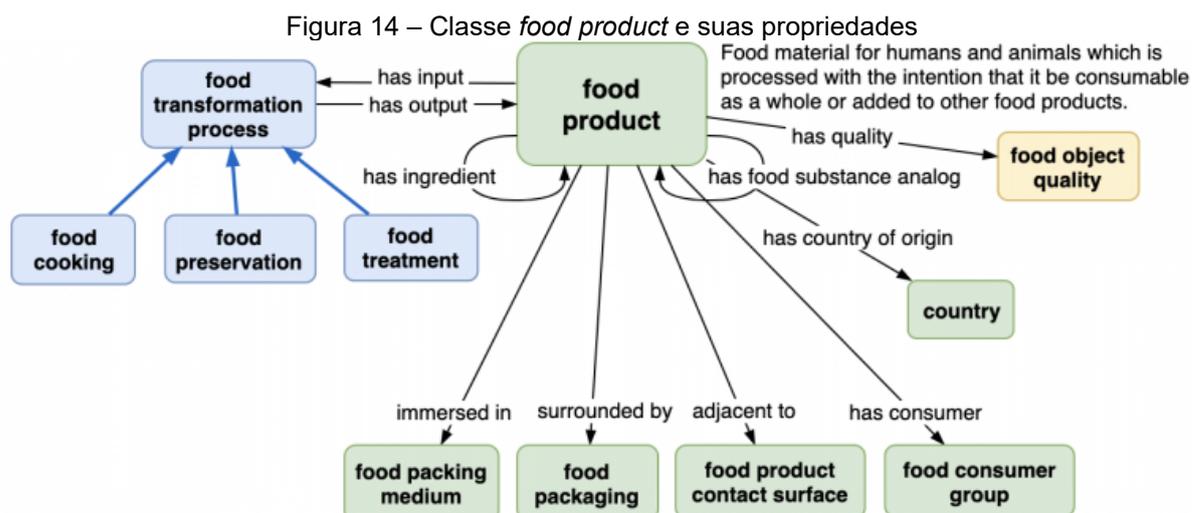
havendo mais um nível de detalhamento após esta subdivisão. Observa-se que o tesauro não apresenta subdivisão para biscoitos salgados.

A estrutura taxonômica do LanguaL pode ser usada de referência para a categorização dos biscoitos, adaptando ao contexto específico, visto que há tipos de biscoitos na estrutura que não são comuns no mercado brasileiro. Além disso, as demais facetas do tesauro podem ser utilizadas para descrição das características de produtos alimentícios.

4.1.4.7 FoodOn

A FoodOn é uma iniciativa mundial com objetivo de desenvolver ontologias públicas sobre alimentos. Pertence ao consórcio de código aberto OBO *Foundry* de ontologias interoperáveis orientadas para ciências da vida e apoia a anotação de dados FAIR, compartilhando objetivos em uma ampla variedade de pesquisas acadêmicas, em setores governamentais e comerciais (FoodOn, 2023). A FoodOn é compatível com a *Basic Formal Ontology* (BFO), ou seja, todas as classes fornecidas pela FoodOn são organizadas a partir de classes da BFO (Top; Timmer; Simsek-Senel, 2020).

No diagrama abaixo, é possível observar a classe de produto alimentício com as suas respectivas propriedades:

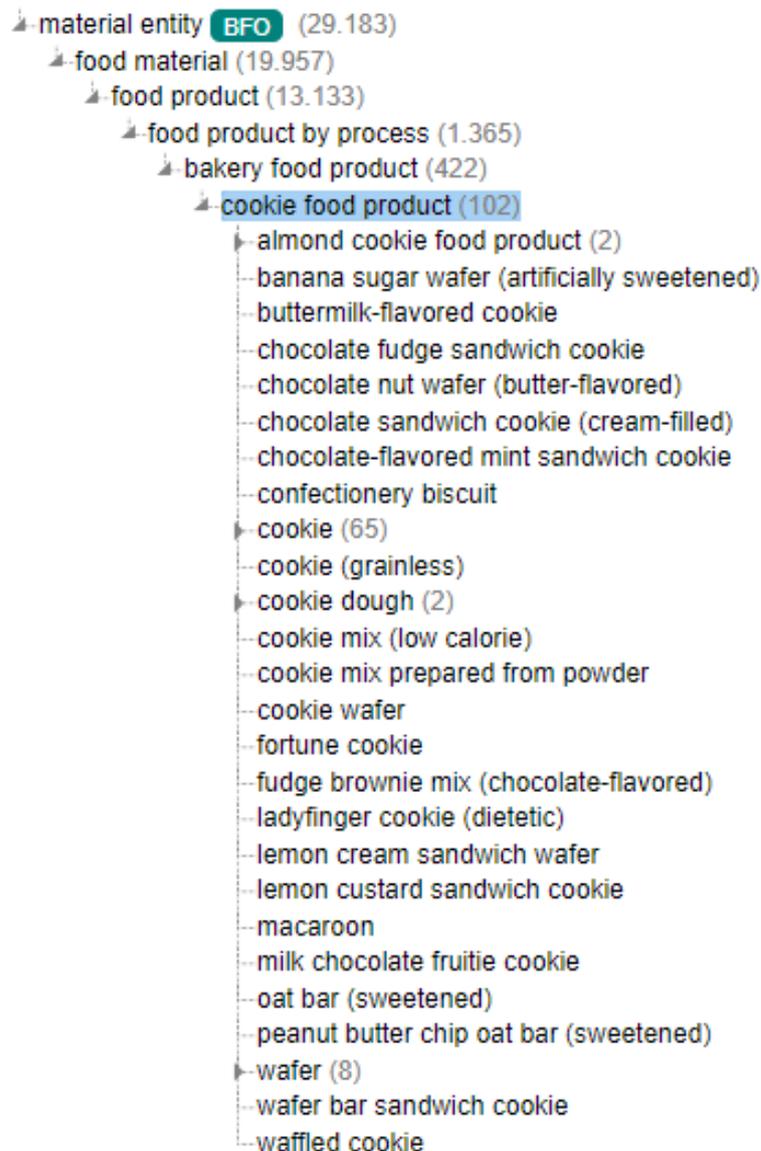


É possível observar algumas semelhanças entre as propriedades da classe de produto alimentício da FoodOn e as facetas do tesauro LanguaL, apresentado anteriormente, como métodos de cozimento, preservação e tratamento, embalagens

e grupos de consumidores. Isso se deve ao fato da FoodOn ser um vocabulário sucessor ao LanguaL (FoodOn, 2023).

Na FoodOn, os biscoitos estão classificados na categoria de produtos alimentícios biscoitos, subordinada à classe de produtos alimentícios de padaria, que por sua vez, encontra-se classificada na classe de produto alimentício por processo. Coordenadas a esta classe, há outras classes de produtos alimentícios, subdivididas a partir de outros critérios de divisão, diferentes de “processo”. Porém, os biscoitos foram localizados na estrutura conforme figura abaixo:

Figura 15 – Classe *Cookie Food Product* na FoodOn



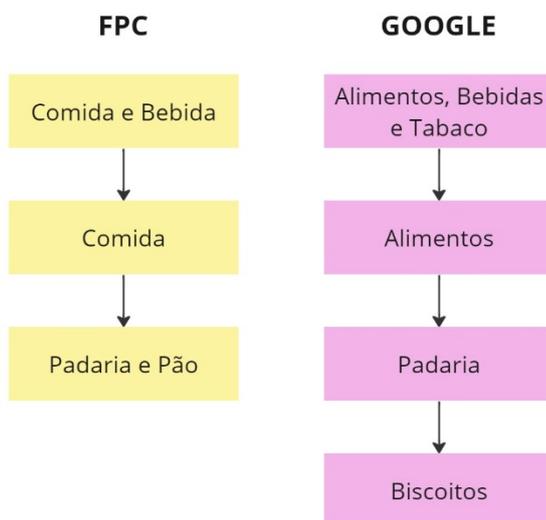
Fonte: FoodOn (2023).

4.1.4.8 Google e Facebook

Além de indicar o Schema.org como padrão de metadados para descrição de produtos, como explanado anteriormente, o Google apresenta uma taxonomia de três a cinco níveis de profundidade para a classificação de itens para venda no Google *Merchant*. Segundo o Google (2024a), todos os produtos recebem uma categoria de taxonomia de forma automática. É possível utilizar o atributo categoria Google do produto [*google_product_category*] para substituir a categorização automática do Google em casos específicos. O Google recomenda utilizar apenas uma categoria, sendo esta a mais específica disponível na estrutura, e que melhor represente o produto que está sendo ofertado.

Já o Facebook (Meta, 2024) oferece uma taxonomia, chamada de FPC, para organizar os itens para venda, em categorias e subcategorias. O Facebook orienta a utilizar a categoria mais específica disponível para classificar os itens que serão vendidos, podendo utilizar uma categoria da FPC ou da taxonomia do Google, apresentada acima. A taxonomia FPC está organizada em três níveis hierárquicos, e se assemelha à estrutura da taxonomia do Google. Na figura abaixo, comparam-se as duas estruturas, referente a categoria de biscoitos.

Figura 16 – Comparação taxonomia FPC e Google



Fonte: Elaborado pela autora.

Na página do Google, há um exemplo de aplicação de categoria para produtos de vestuário, conforme figura abaixo:

Figura 17 – Exemplo de aplicação de categoria do Google



Fonte: Google (2024a).

Como pode-se observar, ambas as taxonomias não especificam os tipos de biscoitos, não sendo suficientes para uma classificação detalhada. Porém, devido à sua relevância no contexto da descrição de produtos na *web*, podem ser usadas como referência, ao menos como uma estrutura inicial, podendo ser incorporadas com outras estruturas mais detalhadas como o GPC, do GS1.

Finalizando esta seção, é possível perceber que não há um consenso quanto à classificação dos biscoitos, tanto da perspectiva de mercado, quanto pelos padrões globais de classificação de produto ou em ontologias especializadas. A legislação também não traz regulamentação da classificação deste tipo de produto alimentício. Porém, analisando os instrumentos apresentados nesta seção é possível observar algumas semelhanças que podem indicar a organização do conhecimento e a terminologia no domínio de análise. Estes aspectos serão utilizados como base para construção do modelo de representação de produtos do varejo nesta pesquisa.

Portanto, serão adotados como referência nesta pesquisa:

- a) Schema.org: serão utilizados os metadados das propriedades das classes “Produto” e “Coisa”, por ser recomendado pelo Google para a descrição de

- produtos na web, além de servir de base para outros instrumentos, tornando-se relevante como referência neste trabalho;
- b) *GS1 Web Vocabulary*: serão utilizados os metadados das propriedades das classes “Produto” e “Produtos alimentares, bebidas e tabaco”, visto que é uma extensão do Schema.org, porém, apresenta uma quantidade maior e mais específica de metadados para o contexto de produtos do varejo;
 - c) GPC: será utilizada a estrutura taxonômica dos tipos de biscoitos, as propriedades e lista de valores destas propriedades, pois detalha a estrutura taxonômica da tipologia de biscoitos enquanto produtos alimentícios, visto que é focado neste domínio, chegando no nível de seus atributos, específicos para cada bloco (*brick*) na taxonomia;
 - d) Google e Facebook: devido a relevância para descrição de produtos vendidos na *web*, serão utilizadas como referências as taxonomias do Google e Facebook, em conjunto com a estrutura do GPC;
 - e) FoodOn: serão utilizadas as propriedades da classe “Produtos alimentícios”, visto que estão associadas aos biscoitos enquanto produtos, mas revela outros aspectos relacionados a processos de transformação do alimento e ao empacotamento que não são tratados pelos instrumentos acima.

4.2 DESENVOLVIMENTO DO MODELO DE REPRESENTAÇÃO

Esta seção descreve o procedimento de análise do corpus de pesquisa, conforme definido anteriormente na metodologia, apresentando informações detalhadas sobre o desenvolvimento do modelo de representação de produtos do varejo, a partir de um exemplo de descrição constante na amostra, ou seja, um indivíduo da amostra.

A análise dos dados do corpus de pesquisa foi realizada de acordo com as etapas abaixo, conforme relatado na seção 3.3.4 Análise dos dados.

- a) extrair termos das descrições dos produtos;
- b) identificar as características dos conceitos selecionados, designados pelos termos, por meio da elaboração de definições;
- c) agrupar os conceitos em categorias com base nas características identificadas;

- d) relacionar as categorias, identificando as relações, inclusive entre classes e suas características;
- e) identificar variações denominativas dos conceitos e definir o termo autorizado;
- f) estruturar padrão de metadados para representar formalmente a estrutura conceitual elaborada;
- g) validar a estrutura perante os conceitos e terminologia identificados no domínio de produtos do varejo, para fins de garantia literária.

Na primeira etapa, foram extraídos os termos das descrições, considerando os seguintes parâmetros:

- a) termos no plural foram passados para a forma singular;
- b) termos no feminino passaram para o gênero masculino;
- c) termos no diminutivo foram passados para a forma regular, sem diminutivo;
- d) foram desconsideradas as preposições e artigos constantes nas definições, exceto em casos em que a preposição compõe um único termo.

No quadro abaixo há uma demonstração deste procedimento aplicado a uma descrição da amostra de pesquisa, selecionada por conveniência.

Quadro 4 – Extração de termos de descrição

Descrição
BISCOITO DOCE RECHEADO NEGRESKO CHOCOLATE RECHEIO BAUNILHA COBERTURA DE CHOCOLATE BRANCO PACOTE 103 G
Termos extraídos
103
BAUNILHA
BISCOITO
CHOCOLATE
CHOCOLATE BRANCO
COBERTURA
DOCE
G
NEGRESKO
PACOTE
RECHEADO
RECHEIO

Fonte: Elaborado pela autora.

Na análise com todas as descrições do corpus de pesquisa, foram identificados 105 termos distintos, a partir das 22 descrições de biscoitos que compõem a amostra. O quadro completo com todos os termos extraídos e suas descrições correspondentes

encontra-se disponível no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11081228> (APÊNDICE B – Termos extraídos das descrições dos produtos).

Para cada termo extraído, foi elaborada definição com objetivo de identificar as principais características dos conceitos denominados por estes termos. No quadro abaixo, é apresentada uma demonstração deste processo, a partir da mesma descrição utilizada no exemplo anterior.

Quadro 5 – Definições dos termos extraídos

Termo	Definição
103	Valor numérico que representa o peso do biscoito medido pela unidade de medida GRAMAS.
BAUNILHA	A baunilha é uma especiaria obtida a partir das vagens secas da orquídea <i>Vanilla planifolia</i> , nativa do México e da América Central. É amplamente utilizada na culinária e na indústria alimentícia como aromatizante devido ao seu sabor doce e floral. No contexto de biscoitos, pode ser adicionado como um ingrediente ou por meio de aromatizantes, extratos ou essências na sua composição, de ambas as formas conferindo sabor doce a massa, recheio ou cobertura do biscoito.
BISCOITO	O biscoito, também conhecido como bolacha, consiste em um produto obtido pela mistura de farinhas, amidos ou féculas com outros ingredientes, submetidos a processos de amassamento e cocção, fermentados ou não, podendo apresentar cobertura, recheio, formato e textura diversos. Estes produtos alimentícios devem apresentar as seguintes denominações de venda: "Biscoito" ou "Bolacha".
CHOCOLATE	Chocolate é um produto obtido a partir da mistura de derivados de cacau (<i>Theobroma cacao</i> L.), massa, pasta ou liquor de cacau, cacau em pó ou manteiga de cacau, com outros ingredientes, sendo constituído de, no mínimo, 25% (vinte e cinco por cento) de sólidos totais de cacau. No contexto de biscoitos, pode ser adicionado como um ingrediente ou por meio de aromatizantes, extratos ou essências na sua composição, de ambas as formas conferindo sabor doce a massa, recheio ou cobertura do biscoito.
CHOCOLATE BRANCO	Chocolate branco é um produto obtido a partir da mistura de manteiga de cacau com outros ingredientes, sendo constituído de, no mínimo, 20% (vinte por cento) de sólidos totais de manteiga de cacau. No contexto de biscoitos, pode ser adicionado como um ingrediente ou por meio de aromatizantes, extratos ou essências na sua composição, de ambas as formas conferindo sabor doce a massa, recheio ou cobertura do biscoito.

COBERTURA	Preparação que é colocada sobre um alimento, normalmente utilizada para cobrir massas, biscoitos e outros produtos culinários. A cobertura pode cobrir, revestir ou dar acabamento em um determinado alimento. Pode ser de diversos sabores e texturas.
DOCE	Um dos cinco sabores básicos, dentre azedo, salgado, amargo e umami. O sabor doce geralmente está associado a sensação de prazer ocasionada pela ingestão de açúcares e carboidratos.
G	Abreviatura de GRAMAS, unidade de medida que mede a grandeza de massa.
NEGRESKO	Marca de biscoito do fabricante NESTLÉ BRASIL LTDA.
PACOTE	Tipo de embalagem que pode variar de tamanho, formato e material. Pode se referir a qualquer tipo de embalagem que contenha produtos. Pode se referir a embalagens, de forma genérica.
RECHEADO	Qualidade de alimento que apresenta algum tipo de recheio. No contexto de biscoitos, refere-se ao tipo de biscoito chamado de "biscoito recheado", composto por duas camadas de massa, muitas vezes de textura mais crocante, que são unidas por um recheio. Atualmente existem também biscoitos apenas com uma camada de massa recheada, diferenciando-se de biscoitos com cobertura. Estes casos são categorizados como biscoitos recheados no mercado.
RECHEIO	Preparação que é colocada dentro de um alimento, ou entre camadas, antes ou depois de ser cozido. Pode ser de diversos sabores e texturas.

Fonte: Elaborado pela autora.

O intuito de elaborar as definições nesta etapa da análise é identificar as principais características dos conceitos, para que, a partir disso, seja possível comparar e agrupar conceitos e atributos semelhantes, formando categorias. Portanto, nesta etapa não foi incluído na definição o conceito imediatamente superordenado, pois a estrutura conceitual ainda não está formada nesta etapa. Esta informação será identificada nas próximas etapas do procedimento de análise. Após a formação das categorias e estabelecimento de relações entre elas, o esquema conceitual estará formado e então será possível revisar estas definições, para que estejam de acordo com as orientações indicadas pela norma ISO 704.

Durante o processo de definição dos termos extraídos da amostra, em alguns casos foi necessário recorrer a motores de busca na *internet* para visualizar a imagem da embalagem do produto, a fim de verificar ou confirmar o significado de determinado termo dentro do contexto da descrição. A busca foi realizada a partir da própria descrição ou pelo GTIN do produto.

As definições foram elaboradas a partir de referências que elucidam o significado dos termos, principalmente considerando o domínio em análise. Essas fontes são devidamente citadas junto às definições, as quais estão consolidadas em tabela completa depositada no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11089740> (APÊNDICE C – Definições dos termos extraídos). Para os termos referentes a marcas, submarcas ou linhas de produto, foram utilizados os dados constantes na base da amostra. Nos casos em que não foi possível encontrar definições em referências na literatura, estas foram elaboradas com base na experiência da pesquisadora no domínio de conhecimento.

A etapa de categorização dos conceitos foi realizada com base no processo de conceitualização apresentado pela norma ISO 704 (International Organization for Standardization, 2022) e na metodologia 101 para desenvolvimento de ontologias (Noy; McGuinness, 2001). Estes dois métodos foram utilizados em conjunto de forma adaptada para esta etapa da pesquisa.

Do ponto de vista do trabalho terminológico, a formação de conceitos exige a identificação das propriedades dos objetos. Aqueles que compartilham propriedades semelhantes são agrupados em unidades. Quando esse grupo de objetos é considerado relevante no âmbito do conhecimento, torna-se uma unidade significativa. As propriedades comuns entre os objetos dessa unidade são então abstraídas como características, que são integradas em um conjunto durante o processo de formação do conceito. Portanto, as propriedades identificam objetos no mundo real, que, por sua vez, são abstraídos como conceitos ou categoria; enquanto as propriedades comuns são abstraídas como características fundamentais para a constituição do conceito (International Organization for Standardization, 2022).

Para esta pesquisa, este método foi adaptado, realizando o processo de abstração de todos os termos elencados anteriormente, agrupando-os em categorias, no nível conceitual, a partir de suas similaridades identificadas por meio da análise de suas definições. Nesse sentido, durante esta fase de análise, não se realizou a distinção entre os termos enquanto objetos ou propriedades, considerando-se que essa determinação seria mais apropriada com base nas características dos elementos já agrupados em categorias. Isso permitiria definir se uma categoria se refere a objetos com existência independente ou aos seus atributos. Por isso, para cada categoria foi estabelecido um termo para sua designação, bem como foi elaborada uma definição,

representando a intensão do conceito da categoria, ou seja, a soma total de características que representam o conceito.

O quadro abaixo demonstra o processo de abstração dos termos para categorias conceituais, com as respectivas definições das categorias:

Quadro 6 – Processo de conceitualização

Termo	Categoria	Definição da categoria
103	Peso	Valor numérico que representa o peso do produto. Acompanhado sempre de uma Unidade de Medida.
BAUNILHA	Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
BISCOITO	Tipo de biscoito	Categoria que agrupa os tipos de biscoito, conforme sua denominação de venda no mercado.
CHOCOLATE	Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
CHOCOLATE BRANCO	Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
COBERTURA	Cobertura	Preparação que é colocada sobre um alimento, normalmente utilizada para cobrir massas, biscoitos e outros produtos culinários. A cobertura pode cobrir, revestir ou dar acabamento em um determinado alimento. Pode ser de diversos sabores e texturas.
DOCE	Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se

		refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
G	Unidade de medida	Convenção que acompanha valor numérico representando o peso do produto.
NEGRESKO	Marca	Representação da promessa de valor que uma empresa oferece aos seus clientes, sendo fundamental para estabelecer e manter a imagem, reputação e posicionamento de uma empresa no mercado. Identifica e diferencia os produtos ou serviços de uma empresa dos produtos ou serviços de concorrentes. Também pode transmitir atributos como qualidade, confiabilidade, inovação, tradição, estilo de vida, valores e propósito da empresa.
PACOTE	Embalagem	Tipos de embalagens.
RECHEIO	Recheio	Preparação que é colocada dentro de um alimento, ou entre camadas, antes ou depois de ser cozido. Pode ser de diversos sabores.

Fonte: Elaborado pela autora.

A partir das definições dos termos extraídos das descrições dos produtos, é possível observar que os conceitos “baunilha”, “chocolate” e “chocolate branco” podem se referir a elementos acrescentados como ingredientes no biscoito, conferindo sabor de baunilha etc., ou podem ser usados outros ingredientes artificiais que também conferem sabor ao biscoito. Portanto, considerando que nesta pesquisa não se teve acesso a lista completa de ingredientes, esta informação não pode ser confirmada e estes conceitos foram categorizados na categoria “sabor”, pois, de qualquer forma, estes elementos conferem sabor ao biscoito. Esta questão está relacionada ao cânone da averiguabilidade ou verificabilidade, disposto por Ranganathan (Gomes; Campos, 2019), em que as características selecionadas devem ser definitivas e verificáveis. Para realizar a verificação desta informação, seria necessário acesso aos produtos em si ou a uma base de dados em que conste a lista de ingredientes.

Ao realizar este processo de conceitualização com todos os termos da amostra, algumas categorias foram ajustadas, bem como suas definições, devido a inclusão de elementos em sua extensão, modificando a intensão do conceito da categoria.

Após agrupar os termos em categorias e definir suas características, foi possível identificar se essas categorias representam objetos com existência independente ou se são atributos desses objetos. Esta etapa foi realizada com base na metodologia 101 para desenvolvimento de ontologias (Noy; McGuinness, 2001). Na etapa 3 desta metodologia, são listados os termos que possam ser importantes para a ontologia. Já a etapa 4 tem objetivo de definir as classes e a hierarquia destas classes; e a etapa 5 busca definir as propriedades destas classes. Nesta pesquisa, a metodologia foi adaptada para incluir a fase de agrupamento em categorias dos termos listados na etapa 3, anteriormente explicada, antes de avançar para as etapas 4 e 5 da metodologia 101. Assim, a identificação das classes e propriedades nas etapas 4 e 5 foi conduzida com base nas definições das categorias estabelecidas, e não nas definições dos termos extraídos das descrições dos produtos.

Nesta etapa, portanto, foram diferenciadas as categorias que representam objetos independentes, ou seja, que são consideradas como classes na estrutura conceitual, e as categorias que representam características destas classes, por dependerem de outro objeto para existir.

O quadro abaixo apresenta a relação de categorias com sua caracterização como classes ou características, com as devidas justificativas:

Quadro 7 – Caracterização das categorias

Categoria	Caracterização	Justificativa
Cobertura	Característica	A cobertura cobre alguma coisa, neste caso, um determinado biscoito.
Embalagem	Característica	A embalagem embala um biscoito específico.
Marca	Característica	Marca de um determinado biscoito.
Peso	Característica	Valor relacionado ao peso de um biscoito específico.
Recheio	Característica	O recheio recheia alguma coisa, neste caso, um determinado biscoito.
Sabor	Característica	O sabor doce ou salgado se refere a algum biscoito com esta característica.
Tipo de biscoito	Classe	O biscoito não depende de outros objetos para existir. Há características que dependem dele para existir.

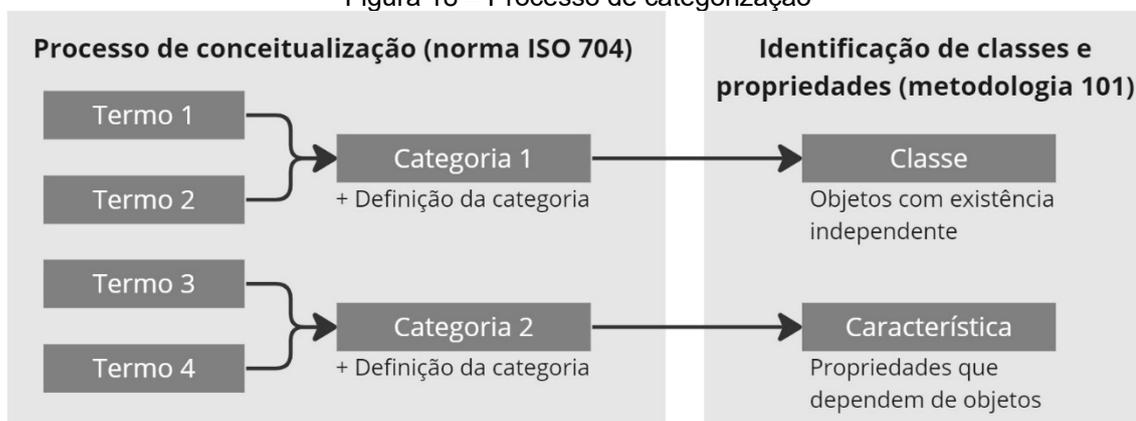
Unidade de medida	Característica	Unidade de medida relacionada ao peso de determinado biscoito.
-------------------	----------------	--

Fonte: Elaborado pela autora.

Na análise com a amostra completa, foram considerados como classes todos os tipos de biscoitos, com objetivo de formar uma estrutura central, hierárquica e tipológica; e foram consideradas características as propriedades referentes a estas classes.

A figura abaixo sumariza o processo de categorização realizado nesta etapa da análise. A categorização completa com todos os dados do corpus de pesquisa será apresentada na próxima seção deste trabalho.

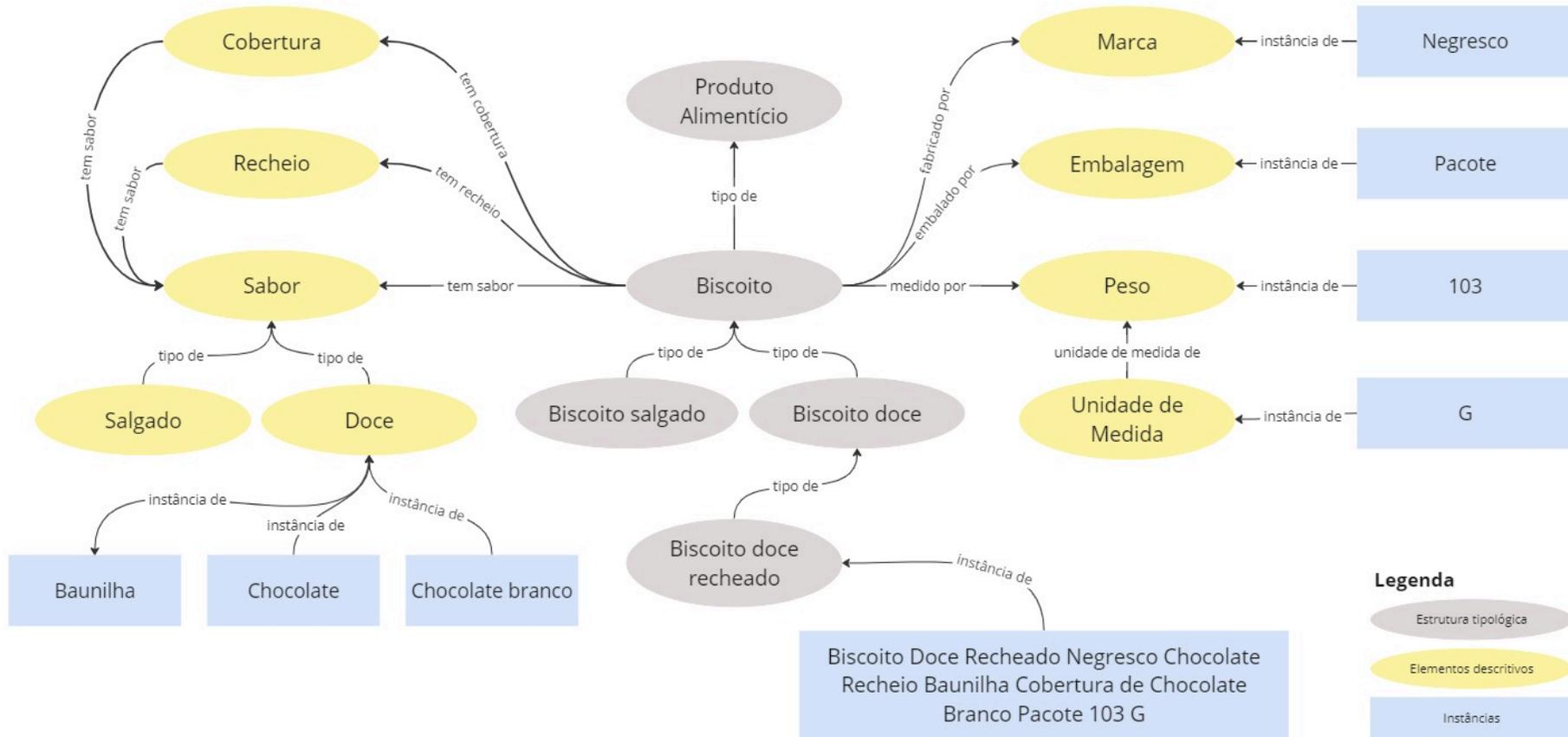
Figura 18 – Processo de categorização



Fonte: Elaborado pela autora.

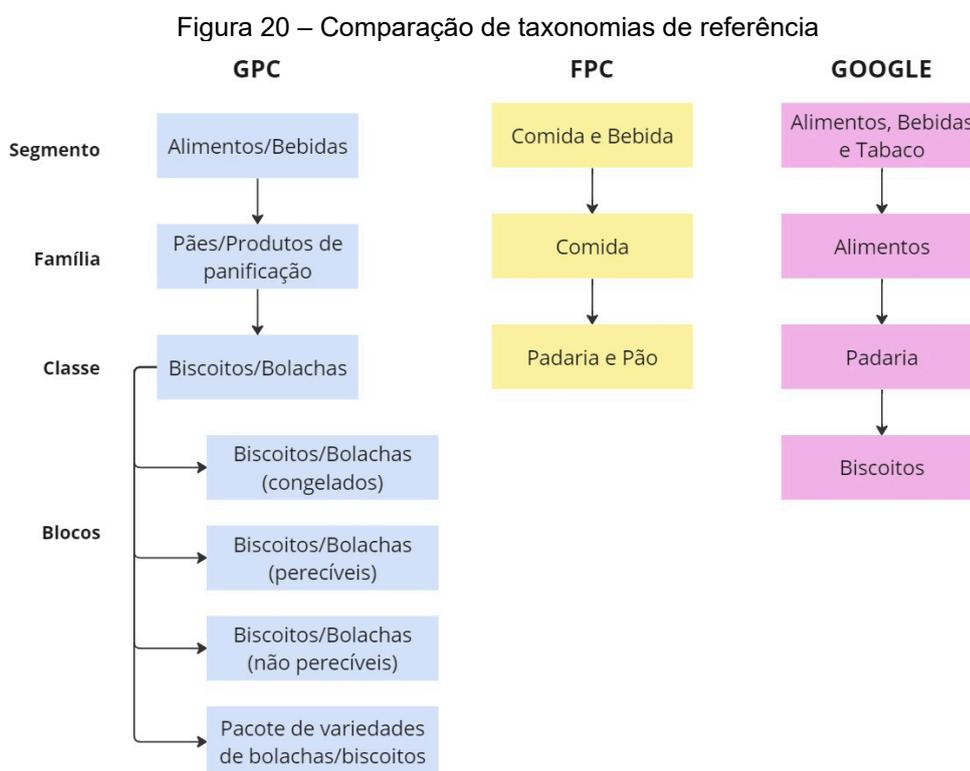
Após identificar as classes e suas características, estas foram relacionadas de forma sistemática, identificando cada tipo de relação, formando um esquema conceitual, conforme figura abaixo:

Figura 19 – Esquema conceitual



Fonte: Elaborado pela autora.

Nas elipses centrais, de cor cinza, encontram-se as classes que representam os tipos de biscoitos, organizadas de forma hierárquica, a partir dos termos extraídos da descrição em análise e de acordo com tipologias de biscoitos encontradas na literatura do domínio de conhecimento. Neste sentido, foram consideradas como referências para esta estruturação as taxonomias do GS1 (GPC), Facebook (FPC) e Google, explicadas anteriormente na seção 4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento. A figura abaixo apresenta uma comparação entre elas, analisando classes relacionadas a biscoitos:



Fonte: Elaborado pela autora.

É possível observar que todas as taxonomias apresentadas iniciam de uma classe que abrange alimentos e/ou bebidas. A partir da definição do conceito de “biscoito” nesta etapa da análise, foi possível identificar que este é um “produto alimentício”, portanto este foi o termo adotado para representar a classe superordenada à classe de biscoitos. Produto alimentício é definido pela Agência Nacional de Vigilância Sanitária (2000) como: “[...] todo alimento derivado de matéria-prima alimentar ou de alimento “in natura”, adicionado, ou não, de outras substâncias permitidas, obtido por processo tecnológico adequado”.

Seguindo com a análise comparativa, a próxima subdivisão abrange produtos de panificação/padaria. Porém, com exceção da classificação NCM, não foi identificado na literatura do domínio, características ou definições associando biscoitos a produtos de panificação. Como as taxonomias do GS1, Facebook e Google foram desenvolvidas em um contexto internacional, é importante levar este aspecto em consideração ao localizar a classificação para o contexto nacional. Portanto a subdivisão para produtos de panificação não foi adotada nesta pesquisa.

A próxima subdivisão do comparativo destina-se a classe de biscoitos, também adotada nesta pesquisa. Após esta classe, apenas a taxonomia GPC apresenta mais subdivisões pelo critério de condições de conservação (congelado, perecível e não perecível), à exceção do pacote de variedades. Na análise de domínio realizada nesta pesquisa, considerando o mercado nacional, este critério não foi identificado como relevante. Na literatura analisada, foi possível perceber que os biscoitos são categorizados de acordo com o seu sabor, como doce ou salgado, e seu tipo, como maria, maisena, rosquinhas etc. Portanto estes foram os critérios de subdivisão adotados nesta pesquisa para a classe de biscoitos.

Como visto anteriormente, a taxonomia de produtos do varejo, chamada de árvore mercadológica, é estruturada de acordo com a árvore de decisão de cada categoria no mercado. A partir da árvore de decisão da categoria de biscoitos, é possível identificar os critérios pelos quais os *shoppers* buscam e escolhem o produto que querem comprar. Estes critérios são organizados de forma hierárquica, de acordo com a importância do critério para decisão do *shopper*. A utilização da árvore de decisão como referência para a escolha do critério de divisão das classes em uma taxonomia respeita o cânone de relevância, disposto por Ranganathan (Gomes; Campos, 2019), que determina que a característica selecionada como critério de divisão deve ser útil à finalidade da classificação.

O SA Varejo (2019) apresenta árvore de decisão para biscoitos, com base em dados do fabricante Mondelez, apresentada na seção 4.1.3 Biscoitos, em que o primeiro critério de decisão para categoria de biscoitos é se este é de sabor doce ou salgado. Portanto, a categoria “biscoito” foi subdividida nas categorias “biscoito doce” e “biscoito salgado”. Como a descrição do produto que está sendo analisado trata-se de um biscoito doce, esta classe foi subdividida em “biscoito doce recheado”, pois esta é a forma comumente utilizada para designar estes tipos de biscoitos, compostos por

duas camadas de massa unidas por um recheio. O biscoito em questão torna-se, então, uma instância desta última classe.

Nas elipses laterais, de cor amarela, encontram-se as características que representam os elementos descritivos dos biscoitos que, na perspectiva da OC, podem ser compreendidos como facetas. As facetas são conjuntos resultantes da aplicação de uma determinada característica, que é definida considerando usuários, objetivos e outros elementos relevantes em determinado contexto (PINHO; VITAL, 2016). Em um sistema de recuperação da informação, cada faceta pode ser tratada como um campo de metadados no sistema para representar e recuperar um objeto informacional. Em um *marketplace*, contexto desta pesquisa, os metadados representando as facetas podem também ser disponibilizados como filtros na pesquisa por produtos.

No domínio de conhecimento em análise, o termo “faceta” não é reconhecido normalmente, sendo utilizado o termo “característica” para se referir aos elementos descritivos do produto. Portanto, neste trabalho foi adotado o termo “característica” para representar os elementos que descrevem os biscoitos. As características estão relacionadas à classe central “biscoitos” visto que representam elementos descritivos de qualquer produto abrangido por esta classe, sendo este salgado ou doce. As relações entre as características e a classe de biscoitos estão identificadas, representando as propriedades destas categorias.

A partir da definição do conceito de “doce” foi possível identificar que este é um dos cinco sabores básicos, dentre eles, salgado, azedo, amargo e umami. Como na literatura não foram identificados os sabores azedo, amargo e umami, se referindo a biscoitos, e também não é um sabor que consta como critério na árvore de decisão desta categoria de produtos, a característica “sabor” foi subdividida apenas em “doce” e “salgado”. A partir do exemplo de descrição analisado, foram instanciados os sabores “baunilha”, “chocolate” e “chocolate branco” na subdivisão da característica “sabor”, designada “doce”.

A característica “peso” apresenta uma outra característica relacionada a ela designada “unidade de medida”. Aqui não há uma relação hierárquica, como relatado no caso acima, pois a unidade de medida não é um tipo de peso, mas trata-se de uma convenção que representa o peso de determinado produto e é acompanhado de um valor numérico. Portanto, este valor numérico foi instanciado na característica “peso” e a abreviatura “G”, representando o termo “gramas”, foi instanciada na característica

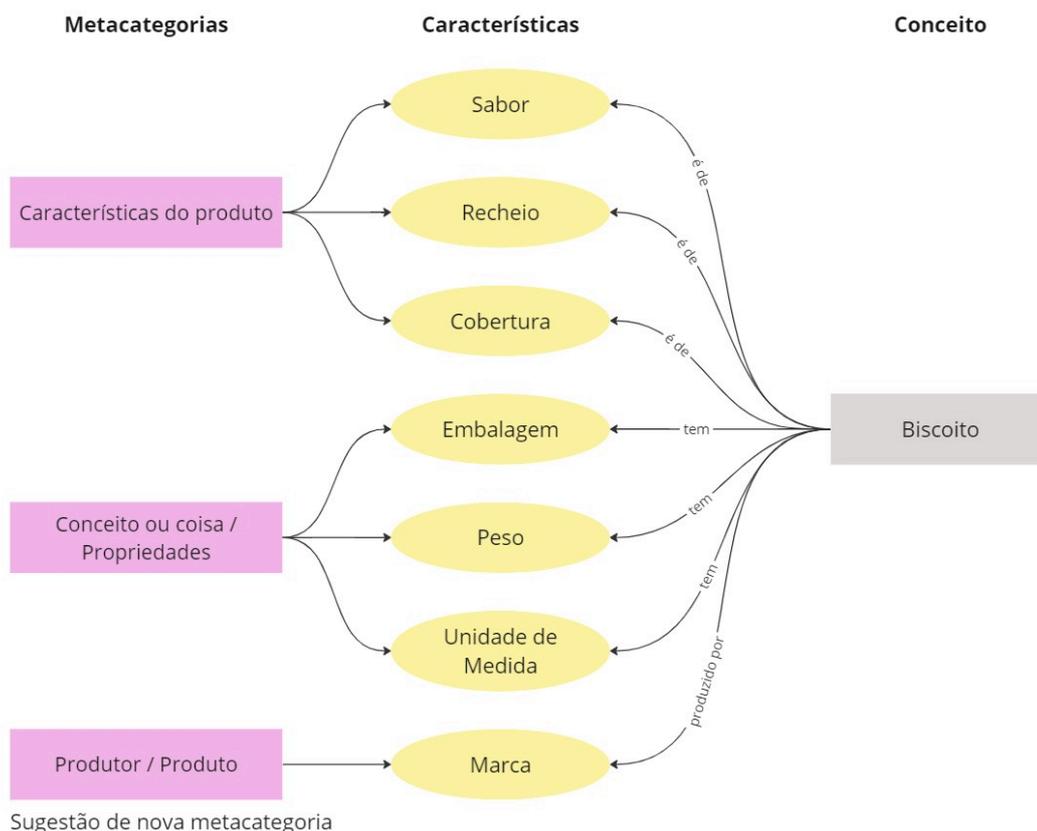
“unidade de medida” e ambas estão relacionadas pela propriedade “unidade de medida de”.

As características “cobertura” e “recheio” estão relacionadas à característica “sabor”, visto que todo recheio ou cobertura apresenta algum sabor, doce ou salgado. No exemplo analisado, o biscoito em si, ou seja, a massa do biscoito, tem sabor de chocolate, porém seu recheio é de sabor baunilha e sua cobertura é de sabor chocolate branco. Portanto, as instâncias da característica “sabor” podem ser utilizadas como instâncias das características “recheio” e “cobertura”, por meio das relações entre estas características, representadas pela propriedade “tem sabor”.

As características que formam o conceito de biscoito podem ser agrupadas em metacategorias, de forma que facilite a aplicação do sistema conceitual em diferentes contextos. Gabriel Junior e Laipelt (2019) apresentam proposta de metacategorias de relações associativas para enriquecimento semântico de tesouros, com base na norma ISO 25964 (International Organization for Standardization, 2011). Dentre as metacategorias propostas, destacam-se as seguintes, para o objetivo desta pesquisa: características do produto (exemplo: “Livro” é de “Papel”), conceito ou coisa/propriedades (exemplo: “Livro” tem no mínimo “50 páginas”) e matéria prima/produto (exemplo: “Aço” é matéria prima de “Estantes”).

A figura abaixo demonstra proposta de aplicação das metacategorias semânticas nas características identificadas:

Figura 21 – Aplicação de metacategorias semânticas



Fonte: Elaborado pela autora.

As características sabor, recheio e cobertura foram categorizadas como “características do produto”, visto que o biscoito é de um determinado sabor, e o recheio e a cobertura também são de um sabor específico. Estas características se referem a materialidade do biscoito enquanto alimento, independentemente da forma pela qual este é comercializado.

Já as características embalagem, peso e unidade de medida foram categorizadas na metacategoria “conceito ou coisa / propriedades”, pois o biscoito, enquanto objeto, apresenta estes atributos em valores específicos. Por exemplo, um determinado pacote (embalagem) tem 100 (peso) gramas (unidade de medida). Estas propriedades se referem a apresentação comercial do produto biscoito. Ou seja, um biscoito com a característica de ser de sabor morango (metacategoria “características do produto”) pode ser comercializado em pacotes de 100 ou 200 gramas (metacategoria “conceito ou coisa / propriedades”).

Não foi identificada metacategoria que abrange a característica “marca”, que está relacionada com a responsabilidade pela ação de produzir determinado biscoito.

Portanto, foi sugerida a criação de nova metacategoria “produtor / produto”, para incluir marcas, submarcas, fabricantes ou indústrias.

Para o modelo de representação final e completo, a extensão das classes foi expandida de acordo com a amostra analisada. Por exemplo, na demonstração realizada nesta seção a categoria “biscoito salgado” não foi subdividida pois a descrição utilizada como exemplo se referia a um biscoito doce. Mas à medida que a amostra de pesquisa apresenta biscoitos salgados, seus elementos descritivos são utilizados na análise para subdivisão das categorias na estrutura conceitual. A estrutura conceitual completa, contemplando a estrutura tipológica e as características, encontra-se na próxima seção deste trabalho.

Na etapa seguinte, foram identificadas variações denominativas dos conceitos e foram definidos os termos autorizados. Do ponto de vista terminológico, a partir da perspectiva da TCT, é fundamental reconhecer e incorporar a variação, contudo, de maneira controlada, a fim de preservar a consistência do sistema conceitual e evitar ambiguidade na recuperação da informação. A variação denominativa é identificada pelas diferentes formas pelas quais pode-se representar um conceito, podendo receber diferentes designações (Freixa, 2002). Em vocabulários controlados, quando há variação denominativa, uma das designações do conceito é escolhida como termo preferido e suas variações são incluídas no sistema como termos equivalentes, possibilitando a recuperação da informação pelas diferentes designações apresentadas para o mesmo conceito (Laipelt; Monteiro-Krebs, 2021).

Para definir quais seriam os termos autorizados e quais seriam as variações, primeiramente foram identificados os elementos da estrutura conceitual que necessitam de controle de vocabulário. De acordo com a ANSI/NISO Z39.19:2005 (National Information Standards Organization, 2010) o controle de vocabulário é usado para melhorar a eficácia dos sistemas de informação, sistemas de navegação na *web* e outros ambientes que têm por objetivo identificar e localizar conteúdos através de sua descrição, utilizando algum tipo de linguagem.

Para a estrutura tipológica, composta por classes de tipos de biscoitos relacionadas hierarquicamente, é necessário que haja controle de vocabulário, pois é preciso estabelecer os termos autorizados que designam as classes. Neste caso, foram identificadas variações terminológicas destes termos com base na própria amostra de descrições de produtos e nas definições elaboradas no processo de análise dos dados.

O quadro abaixo apresenta os termos autorizados, que designam as classes na estrutura tipológica, acompanhados de suas definições, e as variações terminológicas identificadas:

Quadro 8 – Variações terminológicas das classes

Termo autorizado	Definição do termo	Variações
BISCOITO	O biscoito, também conhecido como bolacha, consiste em um produto obtido pela mistura de farinhas, amidos ou féculas com outros ingredientes, submetidos a processos de amassamento e cocção, fermentados ou não, podendo apresentar cobertura, recheio, formato e textura diversos. Estes produtos alimentícios devem apresentar as seguintes denominações de venda: "Biscoito" ou "Bolacha".	Bolacha (variação identificada na definição do conceito de BISCOITO).
BISCOITO DOCE	Tipos de biscoito que apresentam sabor doce em sua massa, recheio ou cobertura. Podem variar em sabor, textura, forma e ingredientes adicionais, como chocolate, frutas, nozes ou especiarias. Eles são frequentemente consumidos como um lanche, acompanhamento de café ou chá, ou como parte de sobremesas.	Bolacha doce.
BISCOITO DOCE RECHEADO	Tipo de biscoito doce composto por duas camadas de massa doce, muitas vezes de textura mais crocante, que são unidas por um recheio doce. Atualmente existem também biscoitos apenas com uma camada de massa recheada, diferenciando-se de biscoitos com cobertura. Estes casos são categorizados como biscoitos recheados no mercado.	Bolacha doce recheada; Biscoito com recheio.
BISCOITO SALGADO	Tipos de biscoito que apresentam sabor salgado em sua massa ou recheio. Podem variar em sabor, textura, forma e ingredientes adicionais, como sal, condimentos ou sabores salgados, como queijo, ervas, alho, cebola, entre outros. Eles são assados até ficarem crocantes e costumam ser consumidos como um lanche salgado, muitas vezes acompanhando queijos, patês, geleias salgadas ou outros aperitivos.	Bolacha salgada.

PRODUTO ALIMENTÍCIO	Produto alimentício é definido pela ANVISA (2000) como: “[...] todo alimento derivado de matéria-prima alimentar ou de alimento "in natura", adicionado, ou não, de outras substâncias permitidas, obtido por processo tecnológico adequado”.	Não identificada.
---------------------	---	-------------------

Fonte: Elaborado pela autora.

No caso das características, foram identificadas três situações em relação aos elementos que fazem parte de cada característica:

- a) valor: valor numérico ou textual em que não é necessário haver controle de vocabulário, portanto não há termo autorizado e variações. Exemplo: “120” na característica “Peso”;
- b) entidade: uma característica enquanto faceta pode apresentar subdivisões. Estas facetas e subfacetas podem ser consideradas entidades no esquema conceitual. Neste caso, assim como ocorre na estrutura taxonômica tipológica central, há necessidade de controle de vocabulário para designação das características e suas subdivisões. Da mesma forma, pode ocorrer controle de vocabulário dos termos das instâncias destas características. Exemplo: característica “Embalagem” e instância “Pacote”;
- c) lista: enumeração de opções padronizadas para uma determinada característica. O termo extraído da descrição pode ser equivalente a alguma das opções da lista. Exemplo: a característica “Alegações nutricionais” (que será apresentada na próxima seção, junto ao modelo de representação completo), traz a subdivisão “Sem glúten”, listando as opções “Sim”, “Não” e “Não informado”. Caso o produto apresente em sua descrição o termo “0% glúten”, será selecionada a opção “Sim” desta característica. Portanto, nestes casos, o controle de vocabulário pode ser realizado através de anéis de sinônimos, visto que não há um termo autorizado, mas sim opções em uma lista de seleção.

O quadro abaixo demonstra a identificação destas situações para as características identificadas no exemplo em análise nesta seção:

Quadro 9 – Controle de vocabulário das características

Característica	Necessidade de controle de vocabulário	Valor / Lista / Entidade
Cobertura	Sim	Entidade
Embalagem	Sim	Entidade

Marca	Sim	Entidade
Peso	Não	Valor
Recheio	Sim	Entidade
Sabor	Sim	Entidade
Unidade de medida	Sim	Lista

Fonte: Elaborado pela autora.

Em resumo, as características que necessitam de controle de vocabulário e seus elementos estão dispostos como entidades, terão um termo autorizado e variações, caso haja. Para os casos de características que necessitam de controle de vocabulário, mas seus elementos estão dispostos como listas, serão elaborados anéis de sinônimos para os termos identificados nas descrições que correspondam às opções enumeradas nas listas. Para as características que trazem um valor “aberto” numérico ou textual não haverá controle de vocabulário.

Comparando estes tipos de características elencados acima com a estrutura de uma ontologia, poderia se dizer que a relação entre as características que contêm entidades e listas, com as respectivas classes, se dá por meio de propriedade de objeto (*domain* e *range*). Já as características do tipo valor são caracterizadas em uma ontologia como propriedade de dados (por exemplo, textual, numérico etc.).

Para os termos referentes a características que necessitam de controle de vocabulário, foram identificadas as suas respectivas variações denominativas. O quadro abaixo demonstra este processo realizado com base no exemplo em análise nesta seção:

Quadro 10 – Variações terminológicas das características

Termo autorizado	Característica	Necessidade de controle de vocabulário	Entidade / Lista	Variações
COBERTURA	Cobertura	Sim	Entidade	
PACOTE	Embalagem	Sim	Entidade	PCT
BAUNILHA	Sabor	Sim	Entidade	
CHOCOLATE	Sabor	Sim	Entidade	
CHOCOLATE BRANCO	Sabor	Sim	Entidade	
NEGRESKO	Marca	Sim	Entidade	
RECHEIO	Recheio	Sim	Entidade	RECHEADO
DOCE	Sabor	Sim	Entidade	
G	Unidade de medida	Sim	Lista	GRAMAS

Fonte: Elaborado pela autora.

Os termos autorizados foram definidos de acordo com os princípios da ISO 704 (International Organization for Standardization, 2022) para criar ou definir novos termos, quando aplicável, sendo eles: princípio da transparência, consistência, adequação, concisão, derivabilidade, composibilidade, correção linguística e preferência pela língua nativa. Estes princípios foram detalhados na seção 2.1.2.3 Terminologia. O quadro completo com todas as variações e os termos autorizados referentes às características encontra-se na próxima seção deste trabalho.

A próxima etapa dos procedimentos metodológicos consiste em estruturar um padrão de metadados para representar formalmente a estrutura conceitual elaborada. Os metadados podem ser considerados como equivalentes a pontos de acesso para recuperação da informação, no contexto da catalogação (Alvarenga, 2001). O desenvolvimento do padrão de metadados nesta pesquisa fundamentou-se em padrões de metadados selecionados como referências, conforme exposto na seção 4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento. Abaixo, estão destacados quais elementos de metadados foram utilizados como referência a partir de cada instrumento, com respectiva justificativa:

- a) Schema.org: foram utilizados os metadados das propriedades das classes “Produto” e “Coisa”, por ser recomendado pelo Google para a descrição de produtos na web, além de servir de base para outros instrumentos, tornando-se relevante como referência neste trabalho;
- b) *GS1 Web Vocabulary*: foram utilizados os metadados das propriedades das classes “Produto” e “Produtos alimentares, bebidas e tabaco”, visto que é uma extensão do Schema.org, porém, apresenta uma quantidade maior e mais específica de metadados para o contexto de produtos do varejo;
- c) GPC: foram utilizadas as propriedades e lista de valores destas propriedades, pois detalha os atributos específicos para o bloco (*brick*) da classe “Biscoitos” na taxonomia;
- d) FoodOn: foram utilizadas as propriedades da classe “Produtos alimentícios”, visto que estão associadas aos biscoitos enquanto produtos, mas revela outros aspectos relacionados a processos de transformação do alimento e ao empacotamento que não são tratados pelos instrumentos acima.

Para facilitar a análise, foi elaborado um compilado das propriedades apresentadas pelos quatro instrumentos citados acima, constando a origem da propriedade, a classe a que esta se refere, o nome da propriedade, o tipo de dado ou

valor esperado e descrição da propriedade (se houver). O compilado encontra-se depositado no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11081234> (APÊNDICE D – Compilado de metadados de referência).

A partir das propriedades reunidas no compilado, foi possível identificar quais delas poderiam ser utilizadas ou adaptadas para representar os elementos descritivos, ou seja, as características, identificadas no exemplo de descrição analisado. O quadro abaixo apresenta os metadados utilizados para o modelo com suas respectivas referências:

Quadro 11 – Padrão de metadados

Metadado	Origem	Classe	Propriedade	Tipo de dado/Valor esperado	Definição
Categoria	Schema.org	Produto	categoria	<i>CategoryCode</i> ou <i>PhysicalActivityCategory</i> ou Texto ou Coisa ou URL	Uma categoria para o item. Sinal de maior ou barras podem ser usados para indicar informalmente uma hierarquia de categorias.
Cobertura	GPC	Biscoitos	tipo de recheio/cobertura	Amêndoas, Queijo, Cereja, Cobertura de chocolate, Combinação de pasta de chocolate/frutas, Recheio de creme, Recheio de frutas, Cobertura glaceada, Presunto, Mel, Geleia, Marzipan, Passas, Sal, Sementes, Especiarias, Açúcar, Baunilha, Wafer, Combinação, Não classificado, Não identificado	Indica, com referência à marca, rotulagem ou embalagem do produto, o termo descritivo usado pelo fabricante do produto para identificar um determinado tipo ou variedade de cobertura/recheio que foi adicionado ao produto.
Código da categoria	GS1 <i>Web Vocabulary</i>	Produto	Código da categoria GPC	<i>xsd:string</i>	Código de 8 dígitos (GPC <i>Brick Value</i>) especificando uma categoria de produto de acordo com o padrão GS1 GPC.
Descrição da categoria	GS1 <i>Web Vocabulary</i>	Produto	Descrição da categoria GPC	<i>rdf:langString</i>	Uma descrição do código que especifica uma categoria de produto de acordo com o padrão GS1 GPC.
Embalagem	GS1 <i>Web Vocabulary</i>	Produto	tem embalagem	<i>gs1:PackagingDetails</i>	Detalhes na embalagem de um produto, incluindo tipo, peso e materiais.
	FoodOn	Produtos alimentícios	cercado por	embalagem de alimentos	
	FoodOn	Produtos alimentícios	imerso em	meio de embalagem de alimentos	

Fabricante	GS1 Web Vocabulary	Produto	tem proprietário de marca	<i>gs1:Organization</i>	O proprietário da marca do produto. A organização responsável por alocar o GTIN ao produto.
GTIN	Schema.org	Produto	gtin	texto ou URL	Um número global de item comercial (<i>Global Trade Item Number - GTIN</i>). Os GTINs identificam itens comerciais, incluindo produtos e serviços, usando códigos de identificação numéricos.
	GS1 Web Vocabulary	Produto	gtin	<i>xsd:string</i>	Um número global de item comercial (<i>Global Trade Item Number - GTIN</i>) é a Chave de Identificação GS1 de 14 dígitos usada para identificar produtos. A chave compreende um Prefixo de Empresa GS1 seguido por um Número de Referência do Item e um Dígito de Verificação.
Marca	Schema.org	Produto	marca	marca ou organização	A(s) marca(s) associada(s) a um produto ou serviço, ou a(s) marca(s) mantida(s) por uma organização ou empresário.
	GS1 Web Vocabulary	Produto	tem marca	<i>gs1:Brand</i>	A marca do produto que aparece na embalagem do consumidor.
Nome	Schema.org	Coisa	nome	texto	O nome do item.
	GS1 Web Vocabulary	Produto	nome do produto	<i>rdf:langString</i>	Breve descrição do produto, amigável ao consumidor, adequada para apresentação compacta.
Peso	Schema.org	Produto	peso	valor quantitativo	O peso do produto ou pessoa.
Recheio	GPC	Biscoitos	tipo de recheio/cobertura	Amêndoas, Queijo, Cereja, Cobertura de chocolate, Combinação de pasta de chocolate/frutas, Recheio de creme, Recheio de frutas, Cobertura glaceada, Presunto,	Indica, com referência à marca, rotulagem ou embalagem do produto, o termo descritivo usado pelo fabricante do produto para identificar um determinado tipo ou variedade de cobertura/recheio que foi adicionado ao produto.

				Mel, Geleia, Marzipan, Passas, Sal, Sementes, Especiarias, Açúcar, Baunilha, Wafer, Combinação, Não classificado, Não identificado	
Sabor	Não localizado nas referências	Não localizado nas referências	é de sabor	instância da característica Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
Unidade de Medida	Não localizado nas referências	Não localizado nas referências	tem unidade de medida	instância da característica Unidade de Medida	Convenção que acompanha valor numérico representando o peso do produto.

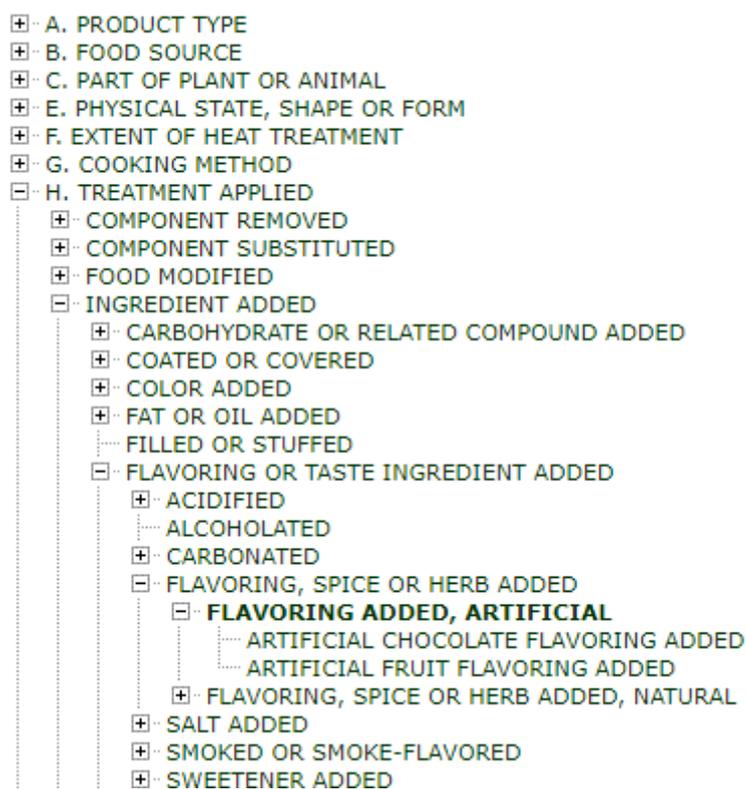
Fonte: Elaborado pela autora.

Os termos utilizados pelos padrões de metadados de referência para designar alguns tipos de dados foram mantidos no idioma original para facilitar a recuperação desta informação nos instrumentos originais. A seleção de metadados elencados no quadro acima considerou os dados obtidos da amostra da pesquisa e também a indicação de dados mínimos necessários para descrição de um produto identificados no domínio de conhecimento. Na perspectiva do Gerenciamento por Categorias, a ECR Brasil (2007) ressalta a importância da padronização da descrição de produtos utilizando pelo menos quatro variáveis: produto (exemplo: amaciante de roupas), marca (exemplo: Comfort), detalhe (exemplo: natural) e conteúdo (exemplo: 500 ml). Já na perspectiva de sistemas de PIM, são determinados quais os atributos serão utilizados como padrão para a representação de determinado tipo de produto (Abraham, 2014). Além disso, são definidas as características de representação de acordo com cada nível da taxonomia, considerando que as classes subordinadas herdam os atributos das respectivas classes superordenadas.

Portanto, foram selecionados metadados que representassem estas informações relacionadas aos biscoitos. A propriedade “nome” associada a classe “Coisa” do Schema.org foi adicionada ao padrão de metadados devido ao Google (2024) indicá-la como propriedade obrigatória para descrição de produtos para vendas *on-line*. A propriedade “GTIN” foi incluída no padrão de metadados pois, apesar de não constar na descrição do produto, é um dado fundamental que está associado ao produto, que o identifica única e globalmente.

Também foram incluídas no padrão, propriedades que não foram identificadas diretamente a partir da descrição exemplificada, mas que estão relacionadas a elas, como fabricante, relacionado à marca do produto; e o código e a descrição da categoria.

A propriedade “sabor” não foi localizada nos padrões de metadados de referência. Pesquisando pelo termo “*flavor*” (“sabor” em inglês) no tesauro facetado LanguaL, descrito na seção 4.1.4, foi localizada a faceta “*Treatment Applied*”, se referindo ao tratamento aplicado a produtos alimentícios. Esta faceta apresenta a seguinte cadeia, do mais geral ao mais específico: ingredientes adicionados > ingrediente de sabor adicionado > aromas, especiarias ou ervas adicionadas > sabor artificial adicionado.

Figura 22 – Estrutura taxonômica da faceta *Treatment Applied*

Fonte: LanguaL (2023).

A propriedade “unidade de medida” também não foi localizada nos padrões de metadados de referência, sendo inserida no padrão a partir da informação referente a “gramas” constante na descrição analisada.

O quadro abaixo demonstra a aplicação do padrão de metadados elaborado, a partir das informações constantes na descrição analisada:

Quadro 12 – Aplicação do padrão de metadados

Propriedade	Valor esperado	Valor
Nome	Texto	Biscoito Doce Recheado Negresco Sabor Chocolate Recheio Baunilha Cobertura Chocolate Branco Pacote 103 G
GTIN	Texto ou URL	7891000352038
Marca	Instância da característica Marca	Negresco
Fabricante	Instância da característica Fabricante	Nestlé Brasil LTDA
Categoria	Categoria da taxonomia	Biscoito Doce Recheado
Sabor	Instância da característica Sabor	Chocolate
Recheio	Instância da característica Sabor	Baunilha
Cobertura	Instância da característica Sabor	Chocolate Branco

Embalagem	Instância da característica Embalagem	Pacote
Peso	Numérico	103
Unidade de Medida	Opção de lista da característica Unidade de Medida	G

Fonte: Elaborado pela autora.

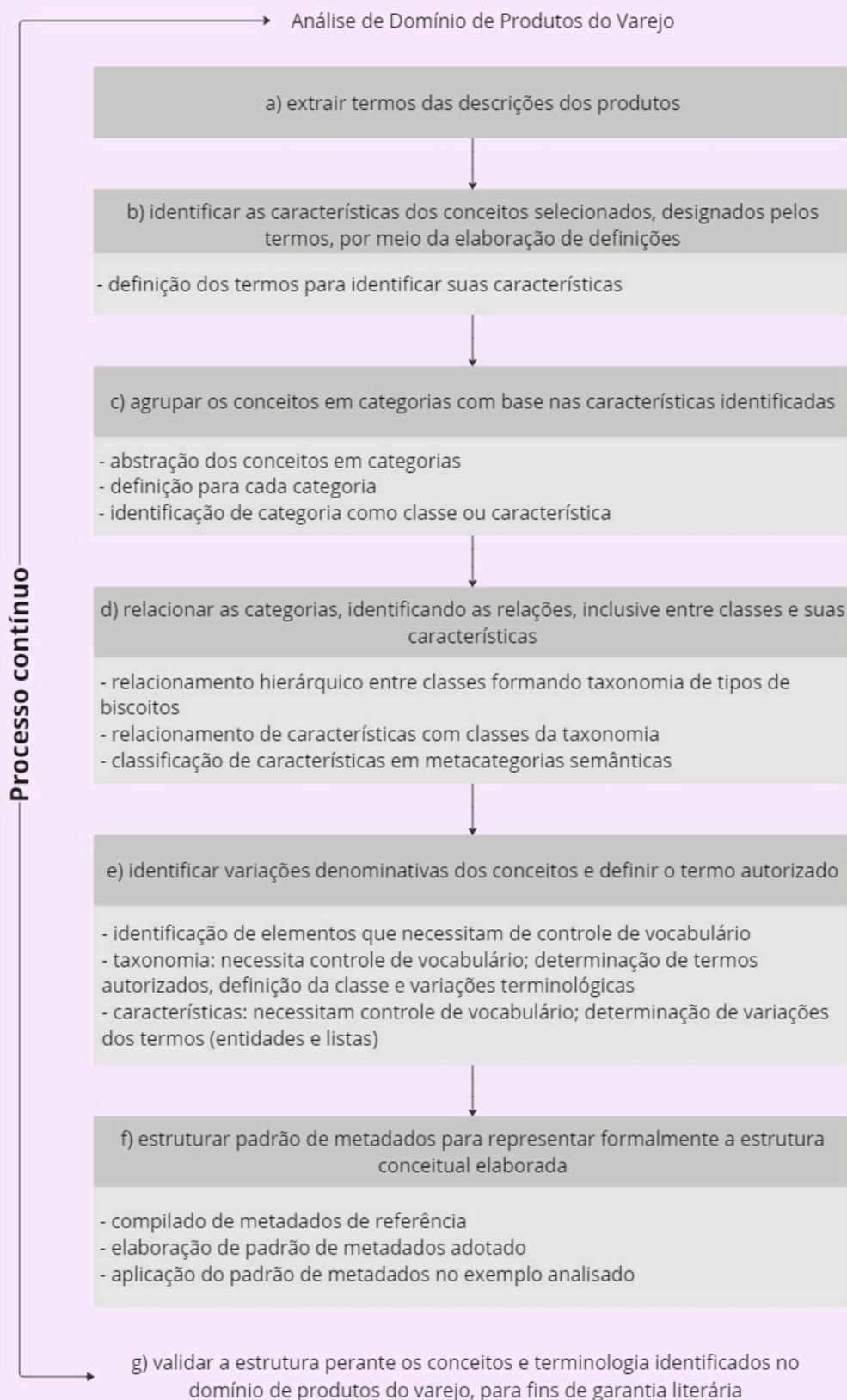
O quadro completo com todos os metadados elencados no modelo, com base na totalidade de descrições da amostra de pesquisa, será apresentado na próxima seção deste trabalho.

O último procedimento metodológico consiste na validação da estrutura perante os conceitos e terminologia identificados no domínio de produtos do varejo, para fins de garantia literária. A validação ocorreu durante cada etapa dos procedimentos metodológicos, à medida que foram elaboradas as definições dos conceitos, a determinação dos termos autorizados e suas variações, definição de categorias e estabelecimento de relações entre elas, buscando a fundamentação teórica na literatura do domínio de conhecimento, já apresentada em seções anteriores.

Finalizando, a figura abaixo sintetiza e esquematiza os procedimentos de análise empregados nesta pesquisa:

Figura 23 – Esquema dos procedimentos metodológicos aplicados na pesquisa

Procedimentos Metodológicos



Fonte: Elaborado pela autora.

Na próxima seção será apresentado o modelo de representação completo, de forma sistemática e formal, com base na totalidade dos dados da amostra de pesquisa.

4.3 MODELO DE REPRESENTAÇÃO DE PRODUTOS DO VAREJO

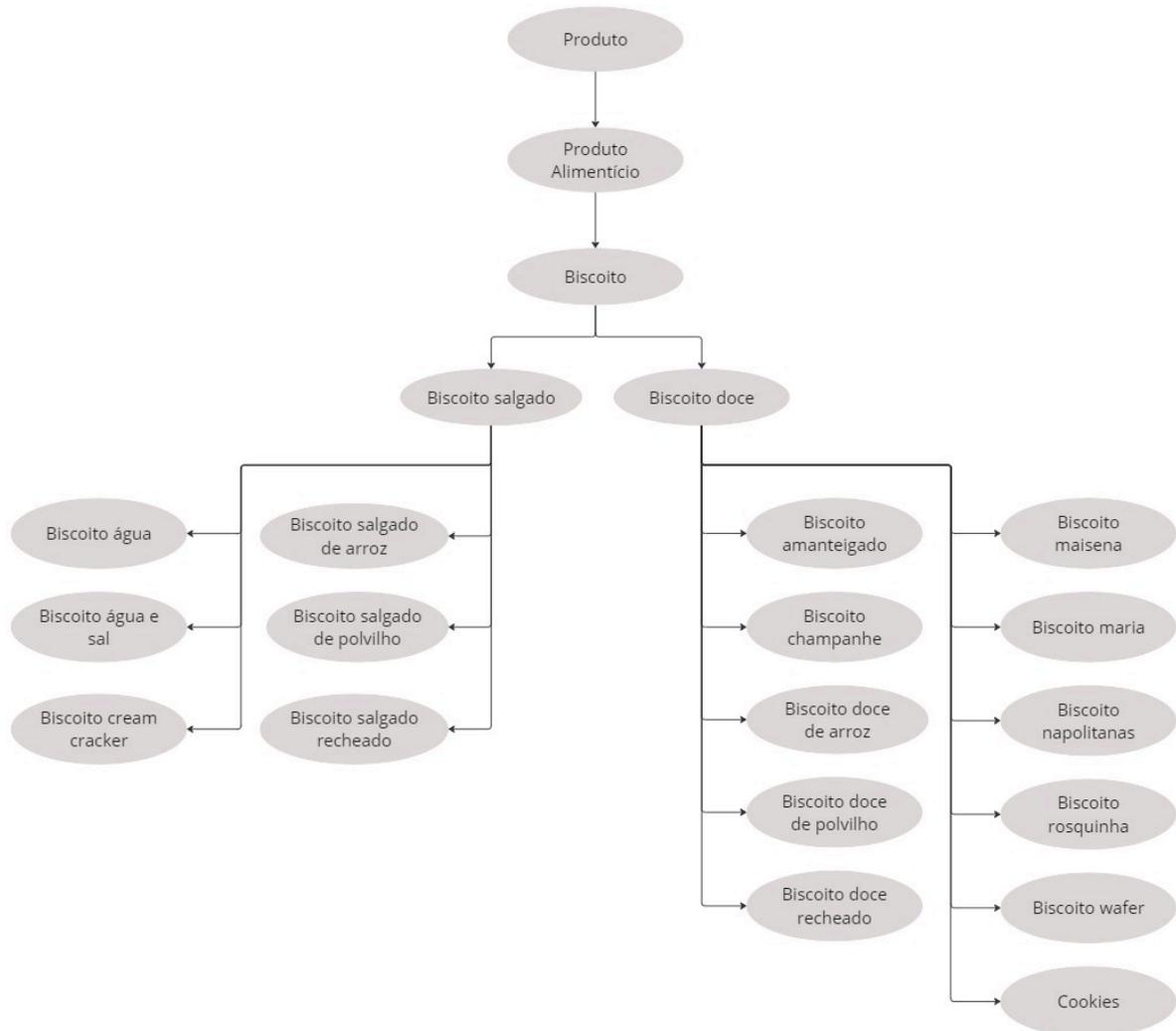
O processo de análise, detalhado na seção anterior, foi executado com todos os dados da amostra, resultando no modelo final de representação de produtos, apresentado nesta seção, objetivando propiciar a aplicação do instrumento para a organização e representação da informação sobre produtos do varejo.

Os resultados serão apresentados em cinco etapas:

- a) apresentação sistemática da taxonomia (tipologia de biscoitos);
- b) apresentação sistemática das características e suas subdivisões, relacionadas às respectivas classes na taxonomia;
- c) apresentação formal do padrão de metadados adotado;
- d) apresentação formal da taxonomia e características em formato de planilha no Excel;
- e) aplicação do modelo (sistemático e formal) em um exemplo de biscoito salgado.

Iniciando pela primeira etapa, segue abaixo figura do esquema conceitual das classes, representando os tipos de biscoitos, organizadas a partir do estabelecimento de relações hierárquicas entre as classes, formando uma taxonomia.

Figura 24 – Taxonomia de tipos de biscoitos



Fonte: Elaborado pela autora.

Semelhante à estrutura já apresentada na seção anterior, a taxonomia aqui é detalhada, subdividindo a classe “biscoito salgado” e “biscoito doce” pelos tipos de biscoitos identificados na amostra, a partir do processo de análise já apresentado na seção anterior. É possível observar que há tipos de biscoitos caracterizados por uma denominação própria, como biscoito maria, biscoito champanhe, biscoito wafer, biscoito cream cracker, cookies; e também há biscoitos designados a partir de seu formato, como biscoito rosquinha, ou a partir de algum ingrediente em sua composição, como biscoito de arroz ou biscoito de polvilho, ou ainda pela sua característica de conter algum recheio, como no caso do biscoito recheado. Isto pode levar a alguma ambiguidade na classificação, pois, por exemplo, pode haver algum biscoito que seja de arroz e que também tenha recheio. Este problema está relacionado ao cânone de exclusividade, disposto por Ranganathan (Gomes; Campos, 2019), que determina que os elementos em um renque devem ser

mutuamente exclusivos. Porém, no domínio de conhecimento em questão, é desta forma que estes biscoitos são usualmente designados e reconhecidos pelos consumidores. Portanto, nestes casos, ao classificar o biscoito é necessário observar qual sua característica predominante, do ponto de vista do mercado. Ou seja, no exemplo do biscoito doce de arroz com recheio, sua característica predominante é ser feito do ingrediente arroz, portanto será classificado na classe “biscoito doce de arroz”, mesmo apresentando recheio, que se trata de uma característica secundária.

Para orientar a atividade de classificação neste sentido, é importante que a definição destas classes leve em consideração estes aspectos, a fim que não haja erros na classificação, como dois biscoitos iguais classificados em classes diferentes. Uma prática aconselhada nestes casos, é a elaboração de notas de escopo, que podem guiar o trabalho de classificação de profissionais não especializados, visando minimizar erros.

Na taxonomia, a classe de produtos alimentícios foi subordinada à classe de produtos, visto que é um tipo de produto. Isto pode fazer sentido quando é realizada a análise apenas de uma categoria, como a categoria de biscoitos analisada neste trabalho. Porém, considerando que em um contexto real de aplicação haverá outros tipos de produtos comercializados no varejo alimentar, a classe de produtos, superordenada na estrutura apresentada, não é relevante pois todos os itens em um mercado são produtos e toda a taxonomia é destinada para a organização destes produtos. Portanto, buscou-se identificar no domínio de que forma a taxonomia, ou estrutura ou árvore mercadológica, é mais comumente estruturada.

De acordo com Pereira e Barros (2017), a estrutura do conjunto dos produtos em um varejo normalmente é organizada nos seguintes cinco níveis hierárquicos: departamento, categoria, subcategoria, segmento e subsegmento. Já para o ECR Brasil (2007), a estrutura é organizada em quatro níveis: categoria, subcategoria, segmento e subsegmento. De acordo com Parente (2000), no GC, a estrutura pode ser dividida em departamentos, categorias, subcategorias, segmentos e subsegmentos. As categorias são formadas por produtos que são complementares ou substitutos. As subcategorias são agrupamentos de produtos com alto grau de substituição, que concorrem entre si. Os segmentos são produtos que atendem preferências específicas de diferentes grupos de clientes. Os subsegmentos são agrupamentos de produtos com maior grau de especificação.

Neste ponto é importante destacar um conceito comum no domínio de conhecimento: os produtos em uma categoria são substituíveis, ou seja, atendem a uma mesma necessidade de consumo ou uso. No GC, para a etapa de definição da categoria, é necessário conhecer o processo decisório do *shopper*. O resultado deste processo configura-se na árvore de decisão de compra (ECR Brasil, 2007). Como já apresentado anteriormente, o principal critério de decisão do *shopper* na categoria de biscoitos é se este é doce ou salgado, portanto, a classe “biscoitos” foi subdividida desta forma. As classes “biscoitos doces” e “biscoitos salgados” foram subdivididas pelo critério de tipologia de biscoitos, conforme reconhecidos no domínio de conhecimento, a partir de suas características predominantes.

Como é possível observar, não há um consenso ou padrão sobre a quantidade de níveis ou de designação das classes na estrutura mercadológica. Para auxiliar na análise, buscou-se como referência a hierarquia de necessidades apresentada por Kotler e Keller (2012). A partir desta hierarquia, foi realizada uma comparação com os níveis e designações mais comuns identificadas na literatura, a fim de traçar um paralelo entre níveis de necessidade de consumo e os níveis para a organização dos produtos no mercado. O quadro abaixo apresenta este comparativo:

Quadro 13 – Comparativo entre hierarquia de necessidades e níveis de estrutura mercadológica

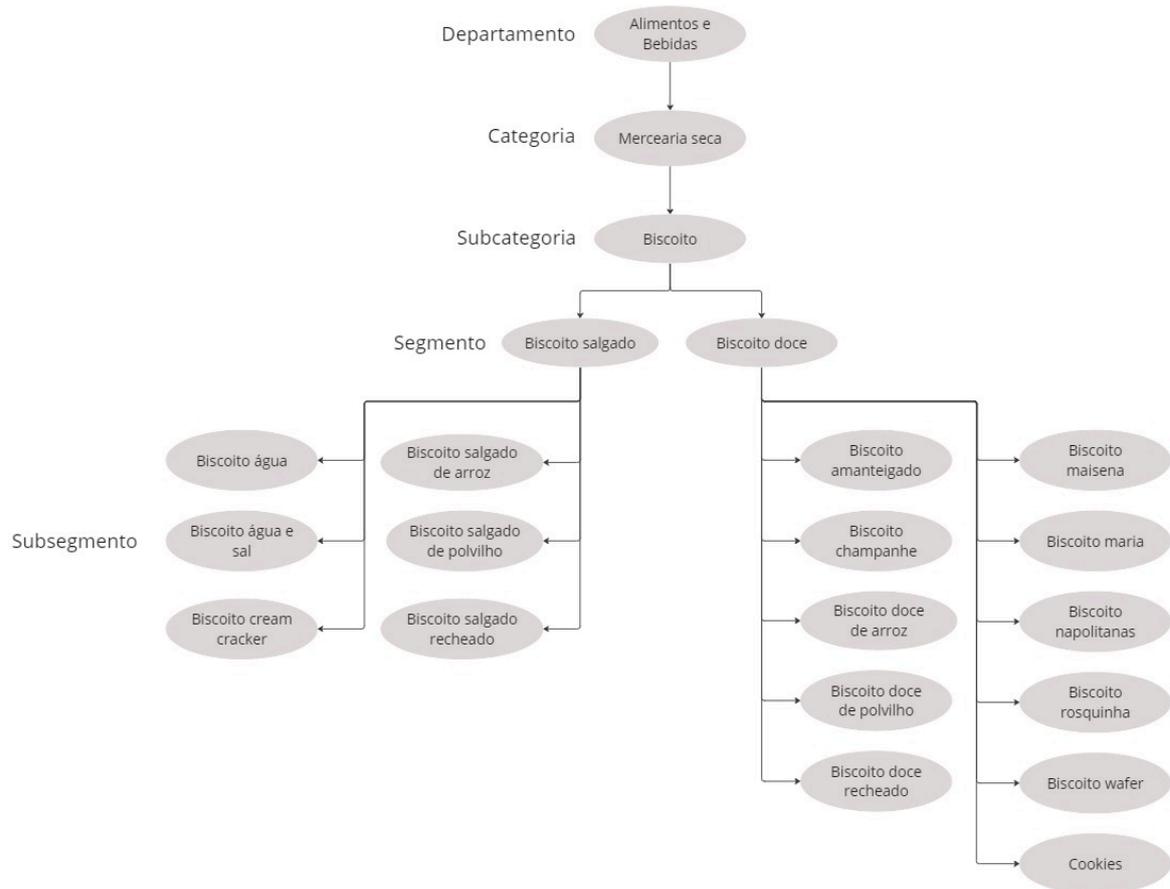
Hierarquia (Kotler; Keller, 2012)	Designação dos níveis da estrutura mercadológica
a) família de necessidade: necessidade central que sustenta a existência de uma família de produtos. Exemplo: alimentos	Departamento
b) família de produtos: todas as classes de produto que atendem a uma necessidade central. Exemplo: alimentos não perecíveis	Categoria
c) classe de produtos: grupo de produtos que compõem uma família de produtos que apresentam uma certa coerência funcional, também conhecido como categoria de produto. Exemplo: biscoitos	Subcategoria
d) linha de produtos: grupo de produtos que compõem uma classe de produtos que apresentam relação por desempenharem função similar, são vendidos para os mesmos grupos de clientes e comercializados pelos mesmos canais. Exemplo: biscoitos doces	Segmento

<p>e) tipo de produto: grupo de itens que compõem uma linha de produtos, compartilhando uma das diversas formas possíveis do produto. Exemplo: biscoito doce recheado</p>	<p>Subsegmento</p>
---	--------------------

Fonte: Elaborado pela autora.

Além disso, conforme explicado anteriormente, foram consideradas como referências para o desenvolvimento da taxonomia, os instrumentos do GS1 (GPC), Facebook (FPC) e Google, explicados anteriormente na seção 4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento. Estas taxonomias, de forma comum, estão estruturadas em quatro níveis: alimentos e bebidas, alimentos, produtos de panificação e biscoitos. A identificação de biscoitos enquanto produtos de panificação não foi identificada nas referências brasileiras no domínio, não se mostrando uma divisão comum. Portanto, com base nos elementos acima relatados, a taxonomia de biscoitos foi reestruturada da seguinte forma:

Figura 25 – Taxonomia final de tipologia de biscoitos



Fonte: Elaborado pela autora.

A categoria foi designada como “mercearia seca”, pois esta é uma diferenciação importante no varejo, especialmente considerando as implicações na cadeia logística para o tratamento de produtos de mercearia, ou seja, não perecíveis, e produtos perecíveis. Portanto, a classe designada “mercearia seca” agrupa outros tipos de produtos, além dos biscoitos, considerados como produtos alimentícios de mercearia seca.

Finalizando esta etapa de apresentação da taxonomia, o quadro abaixo apresenta as definições das classes, juntamente as suas variações terminológicas, caso haja:

Quadro 14 – Definições das classes da taxonomia

Classe	Definição	Variações
Alimentos e Bebidas	Agrupamento de produtos alimentícios perecíveis, não perecíveis e congelados; e bebidas alcoólicas e não alcoólicas.	
Biscoito	O biscoito, também conhecido como bolacha, consiste em um produto obtido pela mistura de farinhas, amidos ou féculas com outros ingredientes, submetidos a processos de amassamento e cocção, fermentados ou não, podendo apresentar cobertura, recheio, formato e textura diversos. Estes produtos alimentícios devem apresentar as seguintes denominações de venda: "Biscoito" ou "Bolacha".	BOLACHA (variação identificada na definição do conceito de BISCOITO)
Biscoito Água	Tipo de biscoito salgado caracterizado por apresentar uma textura crocante e por ser fino e plano, e sabor neutro. Normalmente encontrado em formato quadrado.	
Biscoito Água e Sal	Tipo de biscoito salgado de massa fermentada, composto principalmente por farinha, água e gordura, apresentando textura crocante e sabor levemente salgado. Normalmente encontrado em formato quadrado.	
Biscoito amanteigado	Tipo de biscoito doce composto principalmente por manteiga, farinha e açúcar, apresentando textura delicada sendo leve e quebradiço. Pode ter sabor neutro ou pode ser saborizado com baunilha, chocolate entre outros. Pode ser acompanhado de alguma cobertura ou recheio. Esses biscoitos	

	são frequentemente apreciados como um deleite indulgente e são populares em várias culturas ao redor do mundo.	
Biscoito Champanhe	Tipo de biscoito doce leve e macio, em forma de dedo, feitos de uma massa de bolo esponjosa, geralmente com ovos, farinha, açúcar e às vezes amido de milho. Eles são cozidos até ficarem levemente dourados e têm uma textura delicada e macia. Esses biscoitos são comumente usados em sobremesas como tiramisù, charlottes e outras receitas que exigem camadas de biscoitos mergulhados em líquido, como café ou licores. Eles absorvem líquidos facilmente sem se desintegrar completamente, mantendo sua estrutura e textura. Essa característica os torna ideais para uso em sobremesas que requerem camadas bem definidas e que mantêm sua forma quando cortadas.	CHAMPAGNE (termo oculto a partir de experiência de que pode ser escrito desta forma também)
Biscoito Cream Cracker	Tipo de biscoito salgado plano, seco e crocante, de massa fermentada, tradicionalmente feito com farinha de trigo, fermento, sal e gordura vegetal. Esse biscoito é conhecido por sua textura firme e quebradiça, tornando-o ideal para ser consumido como um lanche leve ou como acompanhamento para queijos, patês, geleias e outras coberturas. Os biscoitos cream cracker têm uma superfície geralmente perfurada com pequenos furos, que ajudam a evitar a formação de bolhas de ar durante o cozimento, resultando em um biscoito mais uniforme e crocante. Eles são frequentemente retangulares em forma e têm uma cor dourada depois de assados.	
Biscoito doce	Tipos de biscoito que apresentam sabor doce em sua massa, recheio ou cobertura. Podem variar em sabor, textura, forma e ingredientes adicionais, como chocolate, frutas, nozes ou especiarias. Eles são frequentemente consumidos como um lanche, acompanhamento de café ou chá, ou como parte de sobremesas.	BOLACHA DOCE

Biscoito doce de arroz	<p>Tipo de biscoito doce feito a partir do arroz, um tipo de cereal, que é cozido, prensado em finas camadas e depois assado para criar uma textura crocante, podendo apresentar alguma cobertura ou recheio doce. Geralmente, os biscoitos de arroz são finos e redondos, embora também possam ser encontrados em outras formas, como retangulares ou quadradas. São conhecidos por sua leveza e crocância, e são frequentemente consumidos como lanches saudáveis, pois geralmente são baixos em calorias e não contêm gorduras saturadas. Além disso, os biscoitos de arroz podem ser consumidos sozinhos ou acompanhados de outros alimentos, como patês, queijos, pastas de amendoim ou geleias, tornando-os uma opção versátil para lanches e refeições leves.</p>	
Biscoito doce de polvilho	<p>Tipo de biscoito doce feito a partir de polvilho doce, ingrediente derivado da mandioca. É um alimento tradicionalmente consumido em várias regiões do Brasil e é conhecido por sua textura leve, crocante e sabor característico, podendo ser apresentado em diversos formatos. Também conhecido como avoador, biscoito de vento, peta, xiringa ou biscoito voador.</p>	<p>AVOADOR; BISCOITO DE VENTO; PETA; XIRINGA; BISCOITO VOADOR (a partir da definição do termo)</p>
Biscoito doce recheado	<p>Tipo de biscoito doce composto por duas camadas de massa doce, muitas vezes de textura mais crocante, que são unidas por um recheio doce. Atualmente existem também biscoitos apenas com uma camada de massa recheada, diferenciando-se de biscoitos com cobertura. Estes casos são categorizados como biscoitos recheados no mercado.</p>	<p>BOLACHA DOCE RECHEADA; BISCOITO COM RECHEIO</p>

Biscoito Maisena	<p>Tipo de biscoito doce feito principalmente de amido de milho (conhecido como maizena em alguns lugares, devido à marca denominada Maizena), que é uma alternativa à farinha de trigo comum em muitas receitas de biscoitos. Esses biscoitos são geralmente muito leves, delicados e quebradiços. São biscoitos de massa dura, conhecidos como estampados, em que o nome "Maisena" ou o nome da marca está gravado na superfície, com desenhos na beira. Comumente consumido como um lanche leve ou como acompanhamento para chá, café ou outras bebidas, ou utilizado em receitas de sobremesas que levam biscoitos em seu preparo, como pavês.</p>	MAIZENA (a partir da definição do termo)
Biscoito Maria	<p>Tipo popular de biscoito doce, fino e crocante, feito com farinha de trigo, açúcar, óleo e essência de baunilha. São biscoitos de massa dura, conhecidos como estampados, em que o nome "Maria" ou o nome da marca está gravado na superfície, com desenhos na beira. Comumente consumido como um lanche leve ou como acompanhamento para chá, café ou outras bebidas, ou utilizado em receitas de sobremesas que levam biscoitos em seu preparo, como pavês.</p>	
Biscoito Napolitanas	<p>Tipo de biscoito doce espanhol feito com canela e coberto com uma mistura de açúcar e canela polvilhada, conferindo textura crocante. Comumente consumido no café da manhã e lanche, acompanhada de café.</p>	
Biscoito rosquinha	<p>Tipo de biscoito doce em formato de anel ou rosquinha. Eles são frequentemente feitos de uma massa de biscoito doce que é moldada em anéis antes de ser assada. Esses biscoitos podem variar em sabor, textura e cobertura, dependendo da receita específica. Os biscoitos rosquinhas podem ser levemente crocantes por fora e macios por dentro, ou podem ser mais firmes e quebradiços.</p>	

Biscoito salgado	Tipos de biscoito que apresentam sabor salgado em sua massa ou recheio. Podem variar em sabor, textura, forma e ingredientes adicionais, como sal, condimentos ou sabores salgados, como queijo, ervas, alho, cebola, entre outros. Eles são assados até ficarem crocantes e costumam ser consumidos como um lanche salgado, muitas vezes acompanhando queijos, patês, geleias salgadas ou outros aperitivos.	BOLACHA SALGADA
Biscoito salgado de arroz	Tipo de biscoito salgado feito a partir do arroz, um tipo de cereal, que é cozido, prensado em finas camadas e depois assado para criar uma textura crocante. Geralmente, os biscoitos de arroz são finos e redondos, embora também possam ser encontrados em outras formas, como retangulares ou quadradas. São conhecidos por sua leveza e crocância, e são frequentemente consumidos como lanches saudáveis, pois geralmente são baixos em calorias e não contêm gorduras saturadas. Além disso, os biscoitos de arroz podem ser consumidos sozinhos ou acompanhados de outros alimentos, como patês, queijos, pastas de amendoim ou geleias, tornando-os uma opção versátil para lanches e refeições leves.	
Biscoito salgado de polvilho	Tipo de biscoito salgado feito a partir de polvilho azedo, ingrediente derivado da mandioca. É um alimento tradicionalmente consumido em várias regiões do Brasil e é conhecido por sua textura leve, crocante e sabor característico, podendo ser apresentado em diversos formatos. Também conhecido como avoador, biscoito de vento, peta, xiringa ou biscoito voador.	AVOADOR; BISCOITO DE VENTO; PETA; XIRINGA; BISCOITO VOADOR (a partir da definição do termo)
Biscoito salgado recheado	Tipo de biscoito salgado composto por duas camadas de massa salgada ou neutra, muitas vezes de textura mais crocante, que são unidas por um recheio salgado. Atualmente existem também biscoitos apenas com uma	BOLACHA SALGADA RECHEADA; BISCOITO COM RECHEIO

	camada de massa recheada, diferenciando-se de biscoitos com cobertura. Estes casos são categorizados como biscoitos recheados no mercado.	
Biscoito Wafer	Tipo de biscoito doce fino e crocante, caracterizado por ser plano e seco, apresentando uma textura leve e porosa, semelhante a uma folha fina. Ele é composto por várias camadas de massa de biscoito intercaladas com camadas de recheio cremoso, como creme de chocolate, baunilha ou outros sabores. Usualmente são consumidos como lanches ou sobremesas e doces, mas também podem ser utilizados no preparo de sobremesas ou para decorar sorvetes.	
Cookies	Tipo de biscoito doce geralmente maior, mais macio e mais denso do que biscoitos convencionais. Cookies frequentemente incluem pedaços de chocolate, frutas secas, nozes ou outros ingredientes adicionais, proporcionando uma variedade de sabores e texturas. Os cookies são frequentemente assados em forma de pequenas bolas de massa em uma assadeira, e podem ser servidos como lanches, sobremesas ou acompanhamentos para café, chá ou leite. O termo "cookies" também se refere ao termo "biscoitos" nos Estados Unidos.	COOKIES (a partir de experiência de que pode ser escrito desta forma também)
Mercearia Seca	Seção ou área da loja no varejo que se dedica à venda de alimentos não perecíveis, ou seja, alimentos que podem ser armazenados em temperatura ambiente e não requerem refrigeração. Esses produtos incluem uma ampla gama de itens como cereais, massas, arroz, enlatados, temperos, biscoitos, snacks, condimentos, farinhas, açúcares e outros itens não líquidos.	

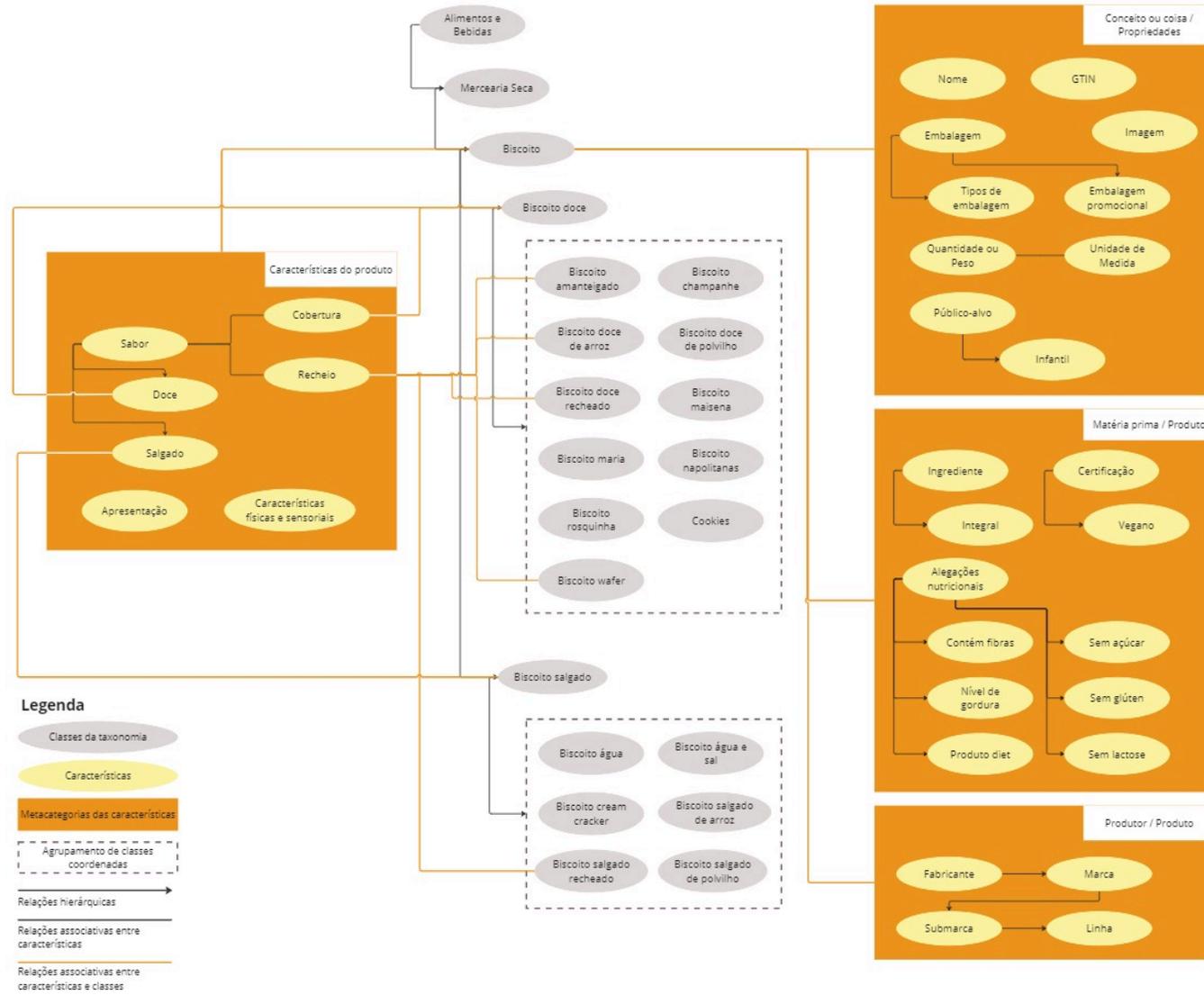
Fonte: Elaborado pela autora.

As definições das classes foram elaboradas de acordo com a norma (International Organization for Standardization, 2022), que determina que a definição

conceitual estabelece um conceito como uma unidade com uma intenção inequívoca, refletida por uma extensão única. Deve incluir o conceito imediatamente superordenado, sendo seguido pelas características delimitadoras. A definição deve ser tão concisa quanto possível e tão complexa quanto necessário, e deve ser formulada de maneira consistente, considerando a linguagem dos usuários e seu nível de conhecimento.

Avançando para a próxima etapa de apresentação do modelo, a figura abaixo apresenta sistematicamente as características e suas subdivisões, relacionadas às suas respectivas classes:

Figura 26 – Apresentação sistemática do modelo de representação



Fonte: Elaborado pela autora.

A forma de apresentação da taxonomia no modelo sistemático foi modificada a fim de facilitar a leitura e compreensão da figura. Portanto, as classes coordenadas, no último nível da taxonomia, foram agrupadas em retângulos pontilhados. Mas as classes e as relações hierárquicas permanecem as mesmas apresentadas anteriormente.

As características foram agrupadas em retângulos de cor laranja de acordo com as respectivas metacategorias, apresentadas na seção 4.2 DESENVOLVIMENTO DO MODELO DE REPRESENTAÇÃO. Devido à forma de apresentação sistemática, a identificação das relações entre características e classes foi demonstrada por meio das metacategorias, considerando suas definições. O quadro abaixo apresenta as metacategorias adotadas nesta pesquisa, com as respectivas definições:

Quadro 15 – Definições das metacategorias

Metacategoria	Definição
Características do produto	Qualidades do biscoito enquanto alimento, independentemente da forma pela qual este é comercializado.
Conceito ou coisa / propriedades	Apresentação comercial do biscoito enquanto produto.
Matéria prima / produto	Elementos que fazem parte da composição do biscoito enquanto alimento.
Produtor / produto	Responsabilidade pela ação de produzir determinado biscoito.

Fonte: Elaborado pela autora.

As metacategorias estão relacionadas à classe de biscoitos, sendo que algumas características nas metacategorias apresentam relações com outras classes específicas na taxonomia. Além disso, algumas características foram subdivididas. A característica “sabor” foi dividida em “doce” e “salgado”, e foram relacionadas às respectivas classes “biscoito doce” e “biscoito salgado”.

As características “cobertura” e “recheio” estão relacionadas à característica “sabor” visto que possuem algum sabor. Portanto, as instâncias na característica “sabor” podem ser utilizadas na própria característica “sabor”, para representar o sabor da massa do biscoito, ou nas características “recheio” ou “cobertura”, para representar o sabor do recheio ou da cobertura, respectivamente. A característica “cobertura” foi relacionada apenas à classe “biscoito doce” pois não foram identificados biscoitos salgados com cobertura tanto nos dados da amostra de pesquisa, quanto no referencial teórico na análise de domínio. Já a característica

“recheio” foi relacionada às classes “biscoito doce recheado” e “biscoito salgado recheado”, devido a ambos possuírem recheio; e também às classes “biscoito amanteigado”, “biscoito doce de arroz” e “biscoito wafer”, pois estes tipos de biscoito podem conter recheio, de acordo com suas definições.

As características “nome”, “GTIN” e “imagem” não tiveram origem a partir da extração dos termos das descrições dos biscoitos, conforme processo de análise relatado. Porém, por se tratarem de informações importantes para a identificação de produto na *web*, foram incluídas como características no modelo. Inclusive, o nome e a imagem são propriedades obrigatórias para oferta de produtos através do Google (2024b).

A característica “embalagem” foi subdividida em “tipos de embalagem” e “embalagem promocional”, visto que são atributos diferentes, mas ambos são aspectos relacionados a embalagem do produto. A característica “unidade de medida” foi relacionada a característica “quantidade ou peso”, pois as instâncias de ambas as características devem ser usadas de forma complementar para descrever a quantidade ou peso de determinado biscoito.

A característica “público-alvo” apresenta uma subdivisão para “infantil”. Esta subdivisão contraria o cânone de diferenciação, disposto por Ranganathan (Gomes; Campos, 2019), que determina que a aplicação de uma característica de divisão deve promover a diferenciação de suas entidades e produzir pelo menos duas classes. Porém, foi mantida no modelo como exemplificação, pois na amostra analisada, foi identificado apenas um biscoito que apresentou informação de se destinar a um público específico. Em uma aplicação real, poderiam ser criadas novas subdivisões, como por exemplo pessoas que praticam atividade física.

A característica “infantil” apresentará uma lista padronizada de opções que podem ser selecionadas de acordo com a situação do produto. O termo que representa esta opção no produto será tratado como uma variação desta opção, em um anel de sinônimos. Na amostra analisada, foi identificado o termo “kids” para se referir ao público infantil. Neste caso, seria escolhida a opção “sim” na lista da característica “infantil” e o termo “kids” seria referenciado como uma variação desta opção em anel de sinônimos. Da mesma forma, a característica “certificação” apresenta apenas a subdivisão para “vegano”, que também apresentará uma lista padronizada de opções.

A característica “ingrediente” pode conter instâncias como entidades, por exemplo arroz para biscoito de arroz, ou tapioca para biscoito de tapioca, mas ela também apresenta uma subdivisão para identificar se seus ingredientes são integrais. A subdivisão “integral” funcionará como lista, conforme mencionado anteriormente.

As alegações nutricionais foram subdivididas em: “contém fibras”, “nível de gordura”, “produto diet”, “sem açúcar”, “sem glúten” e “sem lactose”. Cada uma destas subdivisões apresentará uma lista padronizada de opções que podem ser selecionadas de acordo com a situação do produto, da mesma forma em que ocorre na situação relatada acima. Por exemplo, para um produto em que na embalagem conste o termo “zero açúcar”, na característica “sem açúcar” será selecionada a opção “sim” e o termo “zero açúcar” estará relacionado a opção “sim” por meio de anel de sinônimos.

Por fim, as características “fabricante”, “marca”, “submarca” e “linha” estão relacionadas de forma linear e sequencial, pois cada produto estará associado a uma submarca ou marca específica, assim como a marca deverá estar associada a um fabricante.

O quadro abaixo apresenta as características, suas subdivisões e suas respectivas definições:

Quadro 16 – Definições das características

Característica	Definição
Alegações nutricionais	De acordo com a Agência Nacional de Vigilância Sanitária (2020b), consideram-se alegações nutricionais “[...] qualquer declaração, com exceção da tabela de informação nutricional e da rotulagem nutricional frontal, que indique que um alimento possui propriedades nutricionais positivas relativas ao seu valor energético ou ao conteúdo de nutrientes, contemplando as alegações de conteúdo absoluto e comparativo e de sem adição.”. As propriedades nutricionais podem ser relacionadas a açúcar, lactose, gorduras, colesterol, sódio, sal, ácidos graxos, proteínas, fibras alimentares ou vitaminas e minerais.
Apresentação	Podem ser consideradas Apresentação qualquer característica que diferencie o biscoito de outros do mesmo tipo, que não tenha informação indicativa a qual outra característica se refere, como sabor.
Características Físicas e Sensoriais	Atributos relacionados a aparência e aspectos físicos e sensoriais dos biscoitos.
Certificação	Cerificações conferidas a produtos alimentícios comprovando o uso de determinados ingredientes ou processo de produção.

Cobertura	Preparação que é colocada sobre um alimento, normalmente utilizada para cobrir massas, biscoitos e outros produtos culinários. A cobertura pode cobrir, revestir ou dar acabamento em um determinado alimento. Pode ser de diversos sabores.
Contém fibras	Tipo de alegação nutricional. A fibra dietética ou fibra alimentar é um nutriente presente em alimentos derivados de vegetais. De acordo com a Agência Nacional de Vigilância Sanitária (2020a), a fibra alimentar consta como um dos nutrientes que podem apresentar alegações nutricionais em produtos alimentícios.
Doce	Um dos cinco sabores básicos, dentre azedo, salgado, amargo e umami. O sabor doce geralmente está associado a sensação de prazer ocasionada pela ingestão de açúcares e carboidratos.
Embalagem	Tipos de embalagens e versões promocionais de embalagens no mercado.
Embalagem promocional	Embalagem projetada e usada como parte de uma estratégia de marketing para promover um produto específico. Essas embalagens geralmente são temporárias e são criadas para aumentar o apelo do produto e impulsionar as vendas durante um período específico. Oferecem algum benefício econômico ao <i>shopper</i> , como a mesma quantidade e menor preço, ou maior quantidade pelo mesmo preço, ou ainda alguma parte gratuita do produto.
Fabricante	Empresa responsável pela produção do produto e detentora da marca.
GTIN	Global Trade Item Number (GTIN). Os GTINs identificam itens comerciais, incluindo produtos e serviços, usando códigos de identificação numéricos.
Imagem	Imagem do item, usualmente frontal, mostrando a embalagem do produto.
Infantil	Produtos destinados ao consumo pelo público infantil.
Ingrediente	De acordo com a Agência Nacional de Vigilância Sanitária (2000), considera-se ingrediente "toda substância, incluídos os aditivos alimentares, que se emprega na fabricação ou preparo de alimentos, e que está presente no produto final em sua forma original ou modificada."
Integral	Indica se grãos e cereais da composição do biscoito são integrais. Os grãos e cereais são designados como integrais quando têm todos os seus componentes preservados (farelo, endosperma e germe), mantendo seus nutrientes, fibras e vitaminas. São considerados alimentos integrais quando o produto contiver, no mínimo, 30% de ingredientes integrais e quando a quantidade dos ingredientes integrais for superior à quantidade dos ingredientes refinados.
Linha	Agrupamento de produtos de uma determinada marca ou submarca para públicos específicos ou de acordo com as características do produto.

Marca	Representação da promessa de valor que uma empresa oferece aos seus clientes, sendo fundamental para estabelecer e manter a imagem, reputação e posicionamento de uma empresa no mercado. Identifica e diferencia os produtos ou serviços de uma empresa dos produtos ou serviços de concorrentes. Também pode transmitir atributos como qualidade, confiabilidade, inovação, tradição, estilo de vida, valores e propósito da empresa.
Nível de gordura	Tipo de alegação nutricional. Declaração constante em produtos alimentícios indicando que não há gordura em sua composição. Em relação à gordura trans, o produto pode conter quantidade de gordura trans igual ou menor a 100 (cem) miligramas por 100 (cem) gramas ou mililitros do alimento pronto para o consumo (Agência Nacional de Vigilância Sanitária, 2020a).
Nome	Nome comercial do item.
Produto Diet	Tipo de alegação nutricional. Alimentos diet são produtos isentos de determinado nutriente (como açúcar, glúten, colesterol ou gordura), desenvolvidos para atender a grupos específicos, como as pessoas que sofrem com diabetes ou hipertensão. O termo "diet" pode ser utilizado para os alimentos classificados como alimentos para dietas com restrição de nutrientes, alimentos exclusivamente empregados para controle de peso ou alimentos para dieta de ingestão controlada de açúcares.
Público-alvo	Categoria que agrupa conceitos relacionados a qual público se destina determinado produto.
Quantidade ou peso	Valor numérico que representa a quantidade de unidades de embalagens internas contidas em um produto ou o peso do produto. Acompanhado sempre de uma Unidade de Medida.
Recheio	Preparação que é colocada dentro de um alimento, ou entre camadas, antes ou depois de ser cozido. Pode ser de diversos sabores.
Sabor	Atributo conferido a alimentos por meio da adição de ingredientes naturais ou artificiais, reconhecido pela combinação de dois ou mais sentidos: o paladar (que se refere ao gosto) e o olfato (ao aroma dos alimentos). O sabor é geralmente descrito em termos de doçura, salinidade, acidez, amargor e umami.
Salgado	Um dos cinco sabores básicos, dentre azedo, salgado, amargo e umami. Na culinária, os alimentos salgados são aqueles que têm um sabor salgado pronunciado devido à adição de sal ou a ingredientes naturalmente salgados.
Sem açúcar	Tipo de alegação nutricional. Declaração constante em produtos alimentícios sem açúcar, ou seja, alimentos que contenham no máximo de 0,5 g de sacarose, frutose e ou glicose por 100g ou 100mL do produto final a ser consumido, sem açúcares adicionados e sem outro açúcar declarado com valor diferente de zero (Agência Nacional de Vigilância Sanitária, 2020a).

Sem glúten	Tipo de alegação nutricional. Declaração constante em produtos alimentícios que não contém glúten em sua composição, ou seja, com quantidades não significativas por 100 g ou 100 ml e por porção. O glúten é uma proteína presente no trigo, na cevada, no centeio, na aveia, no triticale, malte e painço, e em todos os seus derivados, como farinha, farelos, germe etc. A declaração que o produto alimentício não contém glúten deve seguir o disposto na legislação sobre rotulagem de alimentos, com objetivo de destacar esta informação aos consumidores que são alérgicos a essa proteína (Agência Nacional de Vigilância Sanitária, 2020a).
Sem lactose	Tipo de alegação nutricional. Declaração constante em produtos alimentícios isentos de lactose, ou seja, alimentos para dietas com restrição de lactose que contêm quantidade de lactose igual ou menor a 100 (cem) miligramas por 100 (cem) gramas ou mililitros do alimento pronto para o consumo, de acordo com as instruções de preparo do fabricante. A lactose é o açúcar presente no leite de todas as espécies de animais mamíferos e seus derivados. A declaração "sem lactose" em produtos alimentícios deve seguir o disposto na legislação sobre rotulagem de alimentos, com objetivo de destacar esta informação aos consumidores que são alérgicos a esse carboidrato (Agência Nacional de Vigilância Sanitária, 2020a).
Submarca	É uma marca que está diretamente ligada a uma marca principal, mas tem sua própria identidade e oferta de produtos específicos. Ela é desenvolvida dentro de uma estratégia de marca maior, aproveitando a reputação, os recursos e o reconhecimento da marca principal para se estabelecer no mercado.
Tipos de embalagem	Tipo de embalagem que pode variar de material, tamanho e formato.
Unidade de medida	Unidade de medida que acompanha valor numérico representando o peso ou quantidade contida no produto.
Vegano	Produtos que não contenham ingredientes de origem animal e/ou seus derivados, incluindo aditivos e adjuvantes. Esta condição é certificada pelos fabricantes e importadores junto aos órgãos nacionais competentes. No Brasil, a Sociedade Vegetariana Brasileira confere esta certificação para produtos sem ingredientes de origem animal, dos quais a empresa, fabricante ou fornecedora, não testa produto finalizado em animais (Sobre o selo [...], 2023).

Fonte: Elaborado pela autora.

Conforme analisado anteriormente, existem características para as quais seus valores necessitam de controle de vocabulário (entidades ou listas) e outras características que não necessitam de controle de vocabulário (valor). Para as características que necessitam controle de vocabulário, foram identificadas variações terminológicas de termos coletados da amostra de descrições de produtos. O quadro abaixo apresenta estes casos:

Quadro 17 – Variações terminológicas das características

Termo	Característica	Lista / Entidade	Variações
0% gordura trans	Alegações nutricionais / Nível de gordura	Lista	SEM GORDURA para característica "Nível de gordura"
Crocante	Características Físicas e Sensoriais	Entidade	CROCANTES (plural, a partir da descrição do produto)
Diet	Alegações nutricionais / Produto diet	Lista	SIM para característica "Produto diet"
Fibra	Alegações nutricionais / Contém fibras	Lista	SIM para característica "Contém fibras"
Fino	Características Físicas e Sensoriais	Entidade	FININHO; FININHOS (a partir da descrição do produto)
G	Unidade de Medida	Lista	GRAMAS
Grãos de aveia	Ingrediente	Entidade	AVEIA; FLOCOS DE AVEIA (a partir da definição do termo)
Kids	Público-alvo / infantil	Lista	SIM para característica "Infantil". Variações CRIANÇA; INFANTIL
Leve mais pague menos	Embalagem promocional	Entidade	LV MAIS PG MENOS
Pacote	Tipo de embalagem	Entidade	PCT
Panda	Características Físicas e Sensoriais	Entidade	URSO PANDA (a partir da definição do termo)
Pasta de amendoim	Ingrediente	Entidade	MANTEIGA DE AMENDOIM; CREME DE AMENDOIM (a partir da definição do termo)
Recheio	Recheio	Entidade	RECHEADO
Sem glúten	Alegações nutricionais / Sem glúten	Lista	SIM para característica "Sem glúten"
Sem lactose	Alegações nutricionais / Sem lactose	Lista	SIM para característica "Sem lactose". Variação: ALIMENTO ISENTO DE LACTOSE
Tradicional	Apresentação	Entidade	ORIGINAL
UN	Unidade de Medida	Lista	UNIDADE; UNID
Vegano	Certificação / Vegano	Lista	SIM para característica "Vegano"
Zero açúcar	Alegações nutricionais / Sem açúcar	Lista	SIM para característica "Sem açúcar"

Zero gordura trans	Alegações nutricionais / Nível de gordura	Lista	SEM GORDURA para característica "Nível de gordura"
--------------------	---	-------	--

Fonte: Elaborado pela autora.

As características que apresentam subcaracterísticas foram separadas por um sinal de barra, na segunda coluna do quadro acima. Conforme explicado anteriormente, para as características que esperam valores do tipo lista, as variações que correspondem ao mesmo significado da respectiva opção da lista, serão tratadas por meio de anéis de sinônimos.

A terceira etapa desta seção, apresenta formalmente o modelo de representação de produtos do varejo, por meio da adoção de um padrão de metadados descritivos. Para isso, os seguintes instrumentos foram usados como referência: Schema.org, *GS1 Web Vocabulary*, GPC e FoodOn. Cada um deles foi explicado mais detalhadamente na seção 4.1.4 SOCs e Padrões de Metadados no domínio de conhecimento. Foi realizado um compilado com todos os metadados destes quatro instrumentos, totalizando 257 propriedades, disponível no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11081234> (APÊNDICE D – Compilado de metadados de referência).

A partir deste compilado, foram listadas todas as características definidas no modelo e foram então buscadas referências destas características neste compilado. O quadro abaixo demonstra o resultado desta análise, apresentando os metadados adotados no modelo final:

Quadro 18 – Padrão de metadados adotado no modelo

Metacategoria	Metadado	Entidade / Lista / Valor	Valor esperado	Referência (origem)	Referência (classe)	Referência (propriedade)
Características do produto	Apresentação	Entidade	Instância da característica Apresentação	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
	Características Físicas e Sensoriais	Entidade	Instância de Características físicas e sensoriais	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
	Cobertura	Entidade	Instância da característica Sabor	GPC	Biscoitos	Tipo de recheio/cobertura
	Doce	Entidade	Instância da característica Doce (Sabor)	GPC	Biscoitos	Salgado/doce
	Recheio	Entidade	Instância da característica Sabor	GPC	Biscoitos	Tipo de recheio/cobertura
	Sabor	Entidade	Instância da característica Sabor	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
	Salgado	Entidade	Instância da característica Salgado (Sabor)	GPC	Biscoitos	Salgado/doce
Conceito ou coisa / Propriedades	Embalagem (embalagem promocional)	Entidade	Instância da característica Embalagem promocional	GS1 <i>Web Vocabulary</i>	Produto	Informação adicional ao consumidor
	Embalagem (tipo de embalagem)	Entidade	Instância da característica Tipo de embalagem	GS1 <i>Web Vocabulary</i>	Produto	Tem embalagem
				FoodOn	Produtos alimentícios	Cercado por

				FoodOn	Produtos alimentícios	Imerso em
GTIN	Valor	Texto ou URL		Schema.org	Produto	gtin
				GS1 <i>Web Vocabulary</i>	Produto	GTIN
Imagem	Valor	URL		Schema.org	Coisa	imagem
				GS1 <i>Web Vocabulary</i>	Produto	Tem imagem do produto
Infantil	Lista	Sim; Não; Não declarado		Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
Nome	Valor	Texto		Schema.org	Coisa	Nome
				GS1 <i>Web Vocabulary</i>	Produto	Nome do produto
Público-alvo	Lista	Usar opções de subdivisão da característica		Schema.org	Produto	Público
	Lista	Usar opções de subdivisão da característica		GS1 <i>Web Vocabulary</i>	Produto	Tem mercado-alvo
Quantidade ou peso	Valor	Numérico		Schema.org	Produto	Peso
				GS1 <i>Web Vocabulary</i>	Produto	Conteúdo líquido
Unidade de Medida	Lista	G; KG; UM		Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
Matéria prima / Produto	Alegações nutricionais (código)	Lista	Usar opções de subdivisão da característica	GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Código de alegação nutricional
	Alegações nutricionais (nome)	Lista	Usar opções de subdivisão da característica	GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Declaração de alegação nutricional

Certificação	Lista	Usar opções de subdivisão da característica	Schema.org	Produto	Tem certificação
			GS1 <i>Web Vocabulary</i>	Produto	Tem certificação
Contém fibras	Lista	Sim; Não; Não declarado	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
Ingrediente	Entidade	Instância da característica Ingrediente	GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Contém informações sobre ingredientes de alimentos e bebidas
			GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Ingrediente principal
			GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Declaração de ingrediente
			FoodOn	Produtos alimentícios	Tem ingrediente
Integral	Lista	Sim; Não; Não declarado	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
Nível de gordura	Lista	Alto nível; Baixo nível; Sem gordura; Não declarado	GPC	Biscoitos	Alegação de nível de gordura
Produto Diet	Lista	Sim; Não; Não declarado	GPC	Biscoitos	Alegação de diabético
			GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Código de tipo de alérgeno dietético marcado na embalagem

	Sem açúcar	Lista	Sim; Não; Não declarado	GS1 <i>Web Vocabulary</i>	Produtos alimentares, bebidas e tabaco	Quantidade básica de açúcares por nutriente
	Sem glúten	Lista	Sim; Não; Não declarado	GPC	Biscoitos	Alegação de produto sem glúten
	Sem lactose	Lista	Sim; Não; Não declarado	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
	Vegano	Lista	Sim; Não; Não declarado	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
Produtor / Produto	Fabricante	Entidade	Instância da característica Fabricante	GS1 <i>Web Vocabulary</i>	Produto	Tem proprietário de marca
				GS1 <i>Web Vocabulary</i>	Produto	Tem fabricante
				Schema.org	Produto	Fabricante
	Linha	Entidade	Instância da característica Linha	Não localizado nas referências	Não localizado nas referências	Não localizado nas referências
	Marca	Entidade	Instância da característica Marca	Schema.org	Produto	Marca
				GS1 <i>Web Vocabulary</i>	Produto	Tem marca
	Submarca	Entidade	Instância da característica Marca	Schema.org	Produto	Marca
GS1 <i>Web Vocabulary</i>				Produto	Tem marca	
Taxonomia	Classe (código)	Entidade	Código da taxonomia (caso haja)	GS1 <i>Web Vocabulary</i>	Produto	Código da categoria GPC
	Classe (descrição)	Entidade	Descrição da taxonomia (caso haja)	GS1 <i>Web Vocabulary</i>	Produto	Descrição da categoria GPC
	Classe (nome)	Entidade	Classe da taxonomia	Schema.org	Produto	Categoria

Fonte: Elaborado pela autora.

A primeira coluna agrupa as características de acordo com sua metacategoria. Nesta coluna, os metadados relacionados às classes da estrutura conceitual foram agrupados como “Taxonomia”. A segunda coluna apresenta o nome do metadado adotado no modelo de representação, seguindo as designações utilizadas para representar as características. A terceira coluna identifica se o metadado é caracterizado como entidade, lista ou valor. A quarta coluna determina qual o tipo de valor é esperado para determinado metadado. Caso o metadado seja do tipo “valor”, este metadado será um campo aberto no sistema de informação, em que o valor esperado será texto, URL ou numérico. Caso o metadado seja do tipo “lista”, os valores esperados são opções padronizadas em uma lista. Caso o metadado seja do tipo “entidade”, os valores esperados são instâncias de características ou de classes da taxonomia.

As três últimas colunas trazem informações referentes aos instrumentos nos quais foram identificados metadados relacionados às características do modelo conceitual. Os casos em que não foram identificados metadados nas referências foram identificados com “Não localizado nas referências”. Para localizar metadados no compilado foi realizada busca por texto por termos relativos às características do modelo, no idioma original dos instrumentos.

Finalizando a apresentação formal do modelo de representação para produtos do varejo, foi elaborada uma outra proposta de apresentação, em formato de planilha em Excel, contemplando a taxonomia e as características. Esta forma de apresentação é mais usual no domínio de conhecimento. Inclusive o próprio GPC é disponibilizado neste formato pelo GS1. Este modelo encontra-se disponível no repositório digital Zenodo: <https://doi.org/10.5281/zenodo.11081251> (APÊNDICE E – Modelo de representação completo em Excel).

Para concluir a apresentação final do modelo de representação, foi selecionada uma descrição da amostra, por conveniência, a fim de aplicar o modelo demonstrando sua utilização. Foi selecionado um biscoito salgado visto que a análise foi realizada anteriormente com a descrição de um biscoito doce. O quadro abaixo apresenta a descrição do produto selecionado, com os termos extraídos, o processo de conceitualização para classes ou características e indicação da classe ou característica a que se refere:

Quadro 19 – Exemplo de aplicação do modelo: extração dos termos e conceitualização

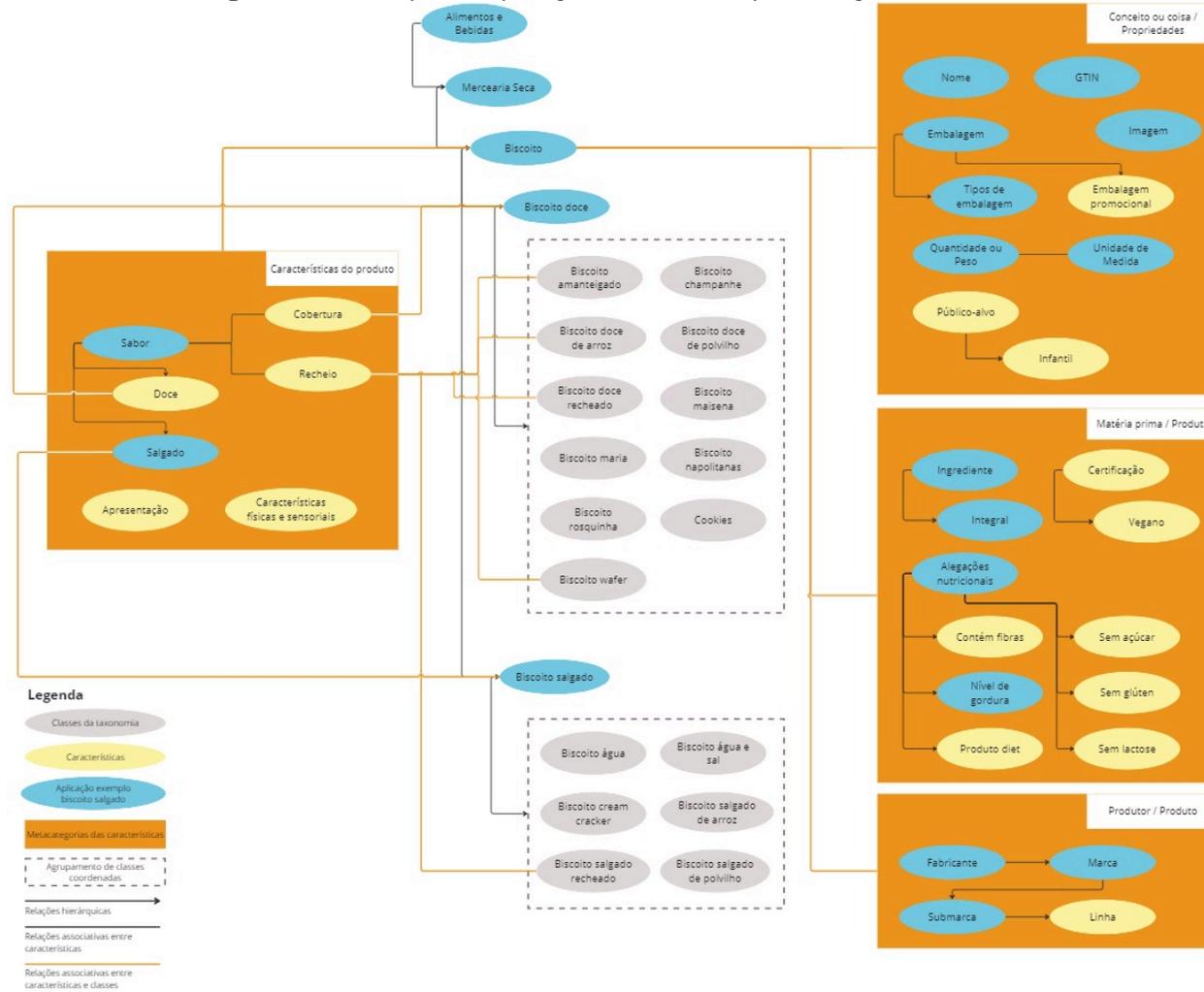
Descrição		
BISCOITO INTEGRAL SALGADO MARILAN LEV COM GRÃOS ANCESTRAIS MANTEIGA COM CHIA E QUINOA 0% GORDURA TRANS PACOTE 170 G		
Termo extraído	Conceitualização	Característica / Classe
170	Característica	Quantidade ou peso
0% GORDURA TRANS	Característica	Alegações nutricionais / Nível de gordura
BISCOITO	Classe	Biscoito salgado
CHIA	Característica	Ingrediente
G	Característica	Unidade de medida
GRÃOS ANCESTRAIS	Característica	Ingrediente
INTEGRAL	Característica	Ingrediente / Integral
LEV	Característica	Submarca
MANTEIGA	Característica	Sabor
MARILAN	Característica	Marca
PACOTE	Característica	Embalagem / Tipo de embalagem
QUINOA	Característica	Ingrediente
SALGADO	Característica	Sabor

Fonte: Elaborado pela autora.

O quadro acima sumariza o processo de análise até então relatado, em que os termos são extraídos da definição, identifica-se o conceito que estes termos designam e realiza-se o processo de conceitualização identificando se o conceito se refere a tipologia do biscoito (classe) ou às suas características. Por fim o biscoito é classificado na taxonomia e são identificadas as características que serão utilizadas para sua representação descritiva.

A figura abaixo apresenta o modelo de representação sistemático, destacando quais classes e características são utilizadas para representação do biscoito do exemplo (elipses destacadas em azul):

Figura 27 – Exemplo de aplicação do modelo: apresentação sistemática



Fonte: Elaborado pela autora.

Pode-se observar que o biscoito não foi classificado no último nível da taxonomia, visto que se trata de um biscoito de sabor salgado, integral, mas sem indicação de um tipo conhecido. Esta situação é comum de ocorrer no domínio de conhecimento, devido à grande variedade de biscoitos lançados no mercado. Além disso, muitas vezes não há informações suficientes sobre o produto para identificar a que tipo se refere. Em sistemas de PIM, é possível permitir que o produto seja classificado apenas no último nível da taxonomia ou em níveis superiores (Abraham, 2014). Portanto, é importante considerar o contexto de negócio para entender se faz mais sentido possibilitar a classificação em níveis superordenados ou criar novas classes subordinadas para classificar novos tipos de produtos.

Por fim, foi elaborada ficha de metadados para o exemplo em questão, com base no padrão de metadados elaborado nesta pesquisa:

Quadro 20 – Exemplo de aplicação do modelo: ficha de metadados

Metadado	Metacategoria	Entidade / Lista / Valor	Valor
GTIN	Conceito ou coisa / Propriedades	Valor	7898091151250
Nome	Conceito ou coisa / Propriedades	Valor	Biscoito salgado integral Marilan Lev com grãos ancestrais sabor manteiga
Imagem	Conceito ou coisa / Propriedades	Valor	---
Embalagem / Tipo de embalagem	Conceito ou coisa / Propriedades	Entidade	Pacote
Quantidade ou peso	Conceito ou coisa / Propriedades	Valor	170
Unidade de Medida	Conceito ou coisa / Propriedades	Lista	G
Sabor / Salgado	Características do produto	Entidade	Manteiga
Ingrediente	Matéria prima / Produto	Entidade	Chia
Ingrediente	Matéria prima / Produto	Entidade	Grãos ancestrais
Ingrediente	Matéria prima / Produto	Entidade	Quinoa
Ingrediente / Integral	Matéria prima / Produto	Lista	Sim

Alegações nutricionais / Nível de gordura	Matéria prima / Produto	Lista	Sem gordura
Fabricante	Produtor / Produto	Entidade	Marilan Alimentos SA
Marca	Produtor / Produto	Entidade	Marilan
Submarca	Produtor / Produto	Entidade	Lev
Classe	Taxonomia	Entidade	Alimentos e Bebidas / Mercearia seca / Biscoito / Biscoito salgado

Fonte: Elaborado pela autora.

O metadado “nome” é destinado para uma descrição aberta, para identificar o produto pelo nome como é descrito comercialmente. Por ser um campo aberto e textual, a forma como esta descrição será elaborada dependerá do contexto de negócio, considerando o público que o varejista atende, bem como questões técnicas de recuperação da informação na *web*. O metadado “imagem” foi mantido na ficha apenas para exemplificação, pois normalmente seria preenchido com a URL de um arquivo da imagem do produto.

O quadro abaixo resume e sistematiza os resultados alcançados com esta pesquisa, demonstrando o atingimento dos objetivos específicos delimitados no início do trabalho:

Quadro 21 – Objetivos específicos e Resultados

Objetivos específicos	Resultados
Identificar e analisar os atributos relevantes dos produtos que compõem o recorte da pesquisa	Identificação e definição das características
Identificar e relacionar categorias de representação a partir destes atributos	Identificação de categorias, enquanto classes e características, e estabelecimento de relações por meio de metacategorias
Estabelecer controle terminológico dos termos adotados nas categorias	Identificação de variações para terminologia adotada nas classes e características
Estruturar modelo de representação para o recorte estudado	Apresentação de modelo de representação de forma sistemática e formal

Fonte: Elaborado pela autora.

Finalizando, esta seção apresentou as versões finais do modelo de representação de produtos do varejo elaborado nesta pesquisa, tanto de forma sistemática, por meio de esquemas conceituais, quanto de maneira formal, por meio de metadados, encerrando com uma demonstração de aplicação em um exemplo de descrição.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo propor um modelo de representação de produtos do varejo, a partir da perspectiva da área de Organização e Representação do Conhecimento e da Informação, no contexto do comércio eletrônico. No decorrer do trabalho, foi trazido referencial teórico na área da Organização e Representação do Conhecimento e da Informação, especialmente em relação a seus instrumentos e teorias que auxiliam o seu desenvolvimento, além de tratar sobre a catalogação enquanto meio para representar elementos descritivos de objetos informacionais, utilizando metadados para formalizar esta representação em sistemas de informação.

Como resultados do trabalho, foi apresentado referencial teórico na área de produtos do varejo, no contexto do comércio eletrônico, como forma de aplicação da análise de domínio enquanto metodologia para desenvolvimento de SOCs; além do próprio modelo de representação para produtos no varejo, com base em análise realizada com amostra constituída em corpus de pesquisa.

Este estudo procurou fornecer contribuições tanto práticas quanto teóricas para os campos da Ciência da Informação (CI) e do varejo. Em relação a contribuições teóricas, destaca-se o esforço deste trabalho em ampliar a compreensão da CI, especialmente no âmbito da OC, ao propor um percurso metodológico para a elaboração de SOCs em um contexto não convencional.

Embora os diferentes tipos de SOCs sejam amplamente definidos na literatura de CI, sua aplicação no contexto do varejo é frequentemente limitada pela falta de familiaridade com esses conceitos. A análise de domínio realizada nesta pesquisa teve como objetivo compreender os principais conceitos da área do varejo e como eles se relacionam, utilizando essa compreensão como base para a modelagem conceitual do modelo de representação. Dessa forma, busca-se aproximar o modelo da realidade e da terminologia da área, facilitando sua adoção no setor varejista.

Neste sentido, a modelagem conceitual desenvolvida nessa pesquisa não se restringiu à criação de um SOC específico, mas se dedicou à compreensão dos elementos fundamentais de cada tipo de SOC e na sua combinação otimizada para representar a realidade dos produtos comercializados no varejo. Essa abordagem visa resolver os desafios de organização e representação da informação enfrentados no setor. Essa perspectiva mais flexível para a elaboração do modelo de representação

promove uma maior convergência entre as duas áreas e facilita a aplicação do modelo no varejo.

A aplicação de um modelo conceitual baseado na perspectiva da OC para a organização e representação de produtos oferece benefícios práticos significativos para o varejo, especialmente no contexto do *e-commerce*, que lida com a representação dos produtos para a sua comercialização em plataformas digitais. Portanto, a organização e o conjunto de informações utilizados para representar estes produtos em um ambiente digital são determinantes para os resultados que os consumidores obtêm ao buscar por esses produtos, influenciando diretamente na recuperação da informação e na decisão de compra. A adoção de ferramentas específicas da OC pode, portanto, aumentar as vendas em *e-commerce*, reduzir o abandono de compras, fidelizar clientes e diminuir o número de devoluções, impactando diretamente o desempenho operacional, comercial e financeiro da empresa. Assim, pesquisas nesse campo apresentam um grande potencial de impacto social e econômico.

Outra contribuição em relação ao referencial teórico na área foi a pesquisa realizada por SOCs e padrões de metadados existentes atualmente no domínio de conhecimento em análise, que resultou em um compilado abrangente e descritivo que pode se tornar referência para futuras pesquisas nesta área.

O desenvolvimento de um modelo de representação de produtos do varejo foi impulsionado por uma motivação prática: ao término da pesquisa, almejava-se que o modelo fosse desenvolvido de maneira a possibilitar seu estudo e aplicação em outras categorias de produtos, bem como em diferentes contextos dentro do comércio eletrônico. Cada fase do procedimento de análise foi detalhada, justificando as decisões tomadas e delineando o percurso metodológico que culminou no modelo final de representação. Ademais, diversas formas de apresentação do modelo foram apresentadas com o intuito de aproximar a aplicação prática do estudo realizado, tornando mais acessível a compreensão das diferentes maneiras de utilizar e aplicar o modelo desenvolvido.

A abordagem metodológica detalhada na pesquisa e as múltiplas formas de apresentação do modelo foram concebidas com o objetivo de oferecer uma compreensão mais profunda de sua aplicação, além de proporcionar um guia claro para sua elaboração e implementação em diferentes contextos do comércio eletrônico. Esses resultados têm o potencial de contribuir tanto para a prática quanto

para o avanço teórico na área, ao fornecer instrumentos tangíveis e aplicáveis que podem aprimorar a organização e representação da informação sobre produtos no comércio eletrônico. Acredita-se que este modelo não apenas melhore as práticas atuais, mas também abra novas possibilidades para futuras pesquisas na interseção entre OC e varejo.

Em relação às limitações da pesquisa, destaca-se a restrição ao acesso a informações primárias sobre os produtos. A falta de acesso aos rótulos e embalagens originais dos produtos, assim como às fichas técnicas das indústrias produtoras, pode haver impactado na profundidade de detalhamento das informações contidas no modelo. No entanto, é válido ressaltar que essa é uma realidade comum na área de produtos do varejo, em que as fontes de dados frequentemente se limitam a planilhas contendo o GTIN e uma descrição genérica do produto. O enriquecimento dessas informações geralmente é realizado pelas equipes de cadastro por meio de pesquisas na *internet* ou consulta a bases de dados pagas.

Este trabalho desenvolveu o modelo de representação com base em apenas uma categoria de produtos do varejo, utilizando uma amostra relativamente pequena. Isso se deve ao foco da pesquisa em uma análise qualitativa, visando analisar e compreender mais profundamente os elementos relevantes para a descrição dos produtos. Dessa forma, é importante reconhecer que o desenvolvimento do modelo com base em amostras de outras categorias de produtos pode revelar elementos que não foram abordados nesta pesquisa. Este aspecto se demonstra como uma possibilidade para estudos futuros.

O modelo foi concebido a partir de uma estrutura central classificatória, conhecida no domínio dos produtos do varejo como uma estrutura ou árvore mercadológica. Esta estrutura mercadológica pode ser utilizada na gestão do varejo, na perspectiva do GC, e serve como base para o menu de navegação de categorias no ambiente digital. No entanto, além dessa estrutura classificatória, é frequentemente necessário atribuir "*tags*" ou palavras-chave aos produtos anunciados para complementar suas descrições e aumentar sua possibilidade de recuperação na busca. Portanto, há oportunidades para estudos futuros no desenvolvimento e aplicação de vocabulários controlados destinados à indexação dos produtos, além de sua classificação.

Este estudo concentrou-se na modelagem de uma estrutura conceitual para representar produtos no varejo, enfatizando os elementos conceituais e descritivos

derivados das informações sobre os produtos. Portanto, abre-se uma oportunidade para futuras pesquisas sobre a aplicação prática deste modelo conceitual em sistemas de informação, visando sua implementação em plataformas de comércio eletrônico. Este aspecto pode ser abordado a partir da perspectiva de PIM, considerando que esta área trata dos processos e ferramentas para gestão de informações sobre produtos, voltando-se principalmente para a aplicação em sistemas de informação.

Ainda na perspectiva tecnológica, aponta-se também oportunidades para estudos e aplicação do modelo de representação conjuntamente com algoritmos de Inteligência Artificial, como modelos para classificação e recomendação, que podem ser usados para automatizar processos no cadastro de produtos, além de melhorar a experiência do usuário.

Além do contexto do varejo e do comércio eletrônico, há oportunidades para expandir esta pesquisa, a partir da perspectiva da Gestão da Informação, com análises sobre a representação de produtos em outros pontos da cadeia de suprimentos, como indústrias produtoras, distribuidores e centros de distribuição. Em cada um desses setores, diferentes informações são relevantes para as suas transações. Ademais, ao considerar a cadeia de suprimentos como um todo, a organização consistente das informações sobre os produtos pode otimizar a eficiência operacional, assegurando a coerência e integridade das informações em todos os elos da cadeia.

Em síntese, este estudo buscou contribuir para o conhecimento sobre a modelagem de uma estrutura conceitual destinada a representação de produtos no âmbito do varejo, considerando sua utilização em plataformas de comércio eletrônico. Ao destacar as oportunidades para pesquisas futuras, se reconhece o potencial prático deste modelo em sistemas de informação, sobretudo no contexto do comércio eletrônico. A integração desses avanços teóricos com a aplicação tecnológica vislumbra possibilidades de melhorias na eficácia e na eficiência das operações no ambiente varejista digital, bem como na experiência do *shopper*.

REFERÊNCIAS

ABRAHAM, Jorij. **Product Information Management: theory and practice**. Cham: Springer International Publishing, 2014.

AGANETTE, Elisangela; ALVARENGA, Lídia; SOUZA, Renato Rocha. Elementos constitutivos do conceito de taxonomia. **Informação & Sociedade**, João Pessoa, v. 20, n. 3, p. 77-93, 2010. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/3994>. Acesso em: 25 abr. 2024.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). **Instrução Normativa nº 75, de 8 de outubro de 2020**. Estabelece os requisitos técnicos para declaração da rotulagem nutricional nos alimentos embalados. Brasília, DF: ANVISA, 2020a. Disponível em: https://antigo.anvisa.gov.br/documents/10181/3882585/IN+75_2020_.pdf/7d74fe2d-e187-4136-9fa2-36a8dcfc0f8f. Acesso em: 14 fev. 2024.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). **Resolução da Diretoria Colegiada nº 711, de 1º de julho de 2022**. Dispõe sobre os requisitos sanitários dos amidos, biscoitos, cereais integrais, cereais processados, farelos, farinhas, farinhas integrais, massas alimentícias e pães. Brasília, DF: ANVISA, 2022. Disponível em: https://antigo.anvisa.gov.br/documents/10181/6482578/RDC_711_2022_.pdf/c739c4a9-6d94-424d-b27b-5ffed15474cf. Acesso em: 25 abr. 2024.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). **Resolução da Diretoria Colegiada nº 429, de 08 de outubro de 2020**. Dispõe sobre a rotulagem nutricional dos alimentos embalados. Brasília, DF: ANVISA, 2020b. Disponível em: https://antigo.anvisa.gov.br/documents/10181/3882585/RDC_429_2020_.pdf/9dc15f3a-db4c-4d3f-90d8-ef4b80537380. Acesso em: 17 fev. 2024.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). **Resolução nº 23, de 15 de março de 2000**. Dispõe sobre O Manual de Procedimentos Básicos para Registro e Dispensa da Obrigatoriedade de Registro de Produtos Pertinentes à Área de Alimentos. Brasília, DF: ANVISA, 2000. Disponível em: https://bvsmms.saude.gov.br/bvs/saudelegis/anvisa/2000/rdc0023_15_03_2000.html. Acesso em: 25 abr. 2024.

ALBERTIN, Alberto Luiz. Pesquisa FGVcia de Comércio Eletrônico no Mercado Brasileiro. **FGV EAESP**, São Paulo, jun. 2020. Disponível em: <https://eaesp.fgv.br/producao-intelectual/pesquisa-fgvcia-comercio-eletronico-mercado-brasileiro>. Acesso em: 19 jul. 2023.

ALBRECHTSEN, Hanne. This is not Domain Analysis. **Knowledge Organization**, [S. l.], v. 42, n. 8, p. 557-561, 2015. Disponível em: https://www.researchgate.net/publication/288642582_This_is_Not_Domain_Analysis. Acesso em: 11 nov. 2022.

ALMEIDA, Mauricio Barcellos. **Ontologia em Ciência da Informação**: teoria e método. Curitiba: CRV, 2020.

ALVARENGA, Lídia. A Teoria do Conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramZero**, Rio de Janeiro, v. 2, n. 6, dez. 2001.

ALVES, Rachel, Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. Marília, 2010. Disponível em: <http://hdl.handle.net/11449/103361>. Acesso em: 25 abr. 2024.

ARAKAKI, Ana Carolina Simionato; ARAKAKI, Felipe Augusto. Dados e metadados: conceitos e relações. **Ciência da Informação**, Brasília, DF, v. 49, n. 3, p. 34-45, set./dez. 2020. Disponível em: <https://doi.org/10.18225/ci.inf.v49i3.5504>. Acesso em: 25 abr. 2024.

ARAKAKI, Felipe Augusto. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família PROV. 2019. 139p. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio Mesquita Filho, Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/180490>. Acesso em: 09 jul. 2023.

ARAKAKI, Felipe Augusto; SIMIONATO, Ana Carolina; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Catalogação e tecnologia: interseções com a web semântica. **Informação@Profissões**, Londrina, v. 6, n. 2, p. 03-19, jul./dez. 2017. Disponível em: <https://doi.org/10.5433/2317-4390.2017v6n2p03>. Acesso em: 25 abr. 2024.

ASHRAF, Jamshaid *et al.* Open eBusiness ontology usage: investigating community implementation of GoodRelations. *In*: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 20., 2011, Hyderabad. **Anais** [...]. Hyderabad: ACM Press, 2011. Disponível em: <http://events.linkedata.org/ldow2011/papers/ldow2011-paper12-ashraf.pdf>. Acesso em: 25 abr. 2024.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE BISCOITOS, MASSAS ALIMENTÍCIAS E PÃES & BOLOS INDUSTRIALIZADOS. **Biscoitos industrializados**: nutrição e indulgência na cultura alimentar. São Paulo: ABIMAPI, 2020.

BARROS, Thiago Henrique Bragato; LAIPELT, Rita do Carmo Ferreira. Uma análise de domínio da área de Organização e Representação do Conhecimento no contexto do periódico Em Questão. **Em Questão**, Porto Alegre, v. 27, n. 4, p. 438-468, out/dez. 2021. Disponível em: <https://doi.org/10.19132/1808-5245274.438-468>. Acesso em: 25 abr. 2024.

BIAGETTI, Maria Teresa. Ontologies as Knowledge Organization Systems. **Knowledge Organization**, [S. l.], v. 48, n. 2, p. 152-176, 2021. Disponível em: https://web.archive.org/web/20220312052742id_/https://www.nomos-elibrary.de/10.5771/0943-7444-2021-2-152.pdf. Acesso em: 25 abr. 2024.

BISCOITO é o “lanche perfeito” para 78% dos trabalhadores. **ABRAS**, 13 jul. 2023. Disponível em: <https://www.abras.com.br/clipping/geral/114356/biscoito-e-o-lanche-perfeito-para-78-dos-trabalhadores>. Acesso em: 03 out. 2023.

BOCCATO, Vera Regina Casari; RAMALHO, Rogério Aparecido Sá; FUJITA, Mariângela Spotti Lopes. A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. **Ibersid**, Zaragoza, v. 2, p. 199-209, 2008. Disponível em: <https://doi.org/10.54886/ibersid.v2i.2235>. Acesso em: 16 dez. 2021.

BÖSE, Dorena. **The development of an approach to online category management with Unilever’s e-commerce drugstore trade partners in Germany**. 2016. Dissertação (Mestrado em Gestão) – NOVA Escola de Negócios e Economia, Carcavelos, 2016. Disponível em: https://run.unl.pt/bitstream/10362/17847/1/Bose_2016.pdf. Acesso em: 25 abr. 2024.

BRÄSCHER, Marisa; CAFÉ, Lígia. Organização da informação ou organização do conhecimento?. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais [...]**. São Paulo: Escola de Comunicação e Artes da USP, 2010. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/176535>. Acesso em: 25 abr. 2024.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 01 mar. 2023.

BUCKLAND, Michael. **Information and Society**. Cambridge: MIT Press, 2017.

CABRÉ, Maria Teresa. **La terminología**: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos. Barcelona: Universitat Pompeu Fabra, 1999a.

CABRÉ, Maria Teresa. **Terminology**: theory, methods and applications. Amsterdam: John Benjamins, 1999b.

CAFÉ, Lígia Maria Arruda; BRATFISCH, Aline. Classificação analítico-sintética: reflexões teóricas e aplicações. **Transinformação**, Campinas, v. 19, n. 3, p. 237-250, set./dez. 2007. Disponível em: <https://www.scielo.br/j/tinf/a/f8FGR3pgqkktTbwgNQgTBVb/?format=pdf&lang=pt>. Acesso em: 25 abr. 2024.

CAMPOS, Maria Luiza de Almeida. Tesouros: principais aspectos teóricos e metodológicos para a sua elaboração. *In*: FUJITA, Mariângela Spotti Lopes; MOREIRA, Walter (org.). **Manual do planejamento, construção e manutenção do Tesouro Unesp para bibliotecas**: do conceitual a práxis. São Paulo: Cultura Acadêmica: 2021. p. 79-93. Disponível em: <https://doi.org/10.36311/2021.978-65-5954-069-3>. Acesso em: 19 dez. 2021.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Taxonomia e Classificação: a categorização como princípio. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais [...]**. Salvador: UFBA, 2007. Disponível em: <http://enancib.ppgci.ufba.br/artigos/GT2--101.pdf>. Acesso em: 25 abr. 2024.

CARLAN, Eliana; MEDEIROS, Marisa Bräscher Basílio. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **RICI: Revista Ibero-americana de Ciência da Informação**, Brasília, DF, v. 4, n. 2, p. 53-73, 2011. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/1675>. Acesso em: 12 dez. 2021.

CATALANI, Luciane *et al.* **E-commerce**. São Paulo: FGV, 2009.

CAVALCANTE, Raphael da Silva; BRÄSCHER, Marisa. Taxonomias navegacionais em sítios de comércio eletrônico: critérios para avaliação. **Transinformação**, Campinas, v. 26, n. 2, p. 191-201, maio/ago. 2014. Disponível em: <https://doi.org/10.1590/0103-37862014000200008>. Acesso em: 25 abr. 2024.

CHAKRABORTY, Goutam.; LALA, Vishal; WARREN, David. An empirical investigation of antecedents of B2B Websites' effectiveness. **Journal of Interactive Marketing**, v. 16, n. 4, p. 51–72, 2002. Disponível em: <https://doi.org/10.1002/dir.10044>. Acesso em: 25 abr. 2024.

CHEN, Donglin; LI, Xiaofei; ZHANG, Jun. User-oriented intelligent service of e-catalog based on semantic web. *In*: IEEE International Conference on Information Management and Engineering, 2., 2010, Chengdu. **Anais [...]**. Chengdu: IACSIT: IEEE, 2010. Disponível em: <https://doi.org/10.1109/ICIME.2010.5477872>. Acesso em: 25 abr. 2024.

CHEUICHE, Paula Veiga *et al.* Aplicação de árvore de decisão para adoção de e-commerce B2B por pontos de venda. **European Journal of Applied Business Management**, [S. l.], v. 5, n. 3, p. 48-67, 2019. Disponível em: <https://nidisag.isag.pt/index.php/IJAM/article/view/452>. Acesso em: 16 nov. 2022.

COYLE, Karen. Library Data in a Modern Context. *In*: COYLE, Karen. **Understanding the semantic web: bibliographic data and metadata**. Chicago: Amer Library Assn Editions, 2010, p. 5-13.

CONWAY, Susan; SLIGAR, Char. Building taxonomies. *In*: CONWAY, Susan; SLIGAR, Char. **Unlocking knowledge assets**. Redmont: Microsoft Press, 2002. cap. 6.

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory for Interconcept. **International Classification**, v. 5, n. 3, p. 142-151, 1978a. Disponível em: https://web.archive.org/web/20220504203022id_/https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142.pdf. Acesso em: 27 abr. 2024.

DAHLBERG, Ingetraut. Knowledge Organization: A New Science? **Knowledge Organization**, [S. l.], v. 33, n. 1, p. 11-19, 2006. Disponível em:

https://www.researchgate.net/publication/288155690_Knowledge_organization_A_new_science. Acesso em: 12 dez. 2021.

DAHLBERG, Ingtraut. Ontology, logic, and Knowledge Organization. **Knowledge Organization**, [S. l.], v. 23, n. 3, p. 129, 1996.

DAHLBERG, Ingtraut. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978b. Disponível em: <http://revista.ibict.br/ciinf/article/view/115>. Acesso em: 12 dez. 2021.

DODEBEI, Vera Lucia Doyle. **Tesauro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Editora Interciência, 2002.

DUBLIN CORE. **DCMI Schemas**. Dublin: DCMI, 2023. Disponível em: <https://www.dublincore.org/schemas/>. Acesso em: 09 jul. 2023.

ECR BRASIL. **Manual de Gerenciamento por categorias**: metodologia simplificada. [S. l.]: ECR Brasil, 2007.

FARINELLI, Fernanda; SOUZA, Amanda D. Ontologias de alto nível: porque precisamos e como usar. **FRC: Fronteiras da Representação do Conhecimento**. Belo Horizonte, v. 1, n. 1, p. 174-202, set. 2021. Disponível em: <https://periodicos.ufmg.br/index.php/fronteiras-rc/article/view/35785>. Acesso em: 27 abr. 2024.

FERNANDES, Pedro; LOPES, Rafael; ZUZUNAGA, Carlos. Plataformas B2B viabilizam maior eficiência para o varejo na América Latina. **McKinsey & Company**, Rio de Janeiro, 6 fev. 2023. Disponível em: <https://www.mckinsey.com.br/our-insights/all-insights/plataformas-b2b-viabilizam-maior-eficiencia-para-o-varejo-na-america-latina>. Acesso em: 11 fev. 2023.

FERREIRA, R. G. O sabor paulista da Piraquê. **IstoÉ Dinheiro**, São Paulo, n. 146, ed. 526, 24 out. 2007. Disponível em: <https://istoedinheiro.com.br/o-sabor-paulista-da-piraque/>. Acesso em: 28 jul. 2023.

FOODON. **Cookie Food Product**. [S. l.]: Ontology Lookup Service, 2023. Disponível em: https://www.ebi.ac.uk/ols4/ontologies/foodon/classes/http%253A%252F%252Fpurl.org%252Fobo%252FFOODON_00001762?lang=en. Acesso em: 19 jul. 2023.

FREIXA, Judit. **La variació terminològica**: anàlisi de la variació denominativa em textos de diferent grau d'especialització de l'àrea de medi ambient. 2002. Tese (Doutorado) – Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, Barcelona, 2002. Disponível em: <http://hdl.handle.net/10803/1677>. Acesso em: 27 abr. 2024.

FREIXA, Judit. Otra vez sobre las causas de la variación denominativa. **Debate Terminológico**, [S. l.], n. 09, p. 38–46, 2013. Disponível em: <https://seer.ufrgs.br/index.php/riterm/article/view/37170>. Acesso em: 27 abr. 2024.

GABRIEL JUNIOR, Rene Faustino; LAIPELT, Rita do Carmo. Descrição das relações semânticas para aplicação em KOS: uso do tesauro semântico aplicado (Thesa). **P2P & Inovação**, Rio de Janeiro, v. 6, n. 1, p. 117-135, 2019. Disponível em: <https://doi.org/10.21721/p2p.2019v6n1.p117-135>. Acesso em: 27 abr. 2024.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GILLILAND-SWETLAND, Ann J. La definición de los metadatos. *In*: BACA, Murtha *et al.* (org.). **Introducción a los metadatos**: vías a la información digital. [S. l.]: Getty, 1999. p. 1-9.

GLYNN, Mark S. How retail category differences moderate retailer perceptions of manufacturer brands. **Australasian Marketing Journal**, v. 15, n. 2, p. 55-67, 2007. Disponível em: [https://doi.org/10.1016/S1441-3582\(07\)70043-5](https://doi.org/10.1016/S1441-3582(07)70043-5). Acesso em: 27 abr. 2024.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida. **A organização do conhecimento na web**: contribuições de Shiyali Ramamrita Ranganathan e Ingetraut Dahlberg. Niterói: IACS/UFF, 2019. (Cadernos Acadêmicos, n. 1).

GOODRELATIONS WIKI. **Documentation/Extensions**. 2023a. Disponível em: <http://wiki.goodrelations-vocabulary.org/Documentation/Extensions>. Acesso em: 09 jul. 2023.

GOODRELATIONS WIKI. **Documentation/Intro**. 2023b. Disponível em: <http://wiki.goodrelations-vocabulary.org/Documentation/Intro>. Acesso em: 09 jul. 2023.

GOODRELATIONS WIKI. **GoodRelations Language Reference**. 2023c. Disponível em: <http://www.heppnetz.de/ontologies/goodrelations/v1.html#ProductOrService>. Acesso em: 09 jul. 2023.

GOOGLE. **Categoria Google do produto [google_product_category]**. 2024a. Disponível em: <https://support.google.com/merchants/answer/6324436#Format&zippy=%2Cprodutos-vendidos-no-google%2Ccampanhas-de-an%C3%BAncios-do-shopping%2Cprodutos-de-vestu%C3%A1rio>. Acesso em: 18 mar. 2024.

GOOGLE. **Dados estruturados do produto**. 2024b. Disponível em: https://developers.google.com/search/docs/appearance/structured-data/product?hl=pt-br#merchant-listings_product. Acesso em: 28 jan. 2024.

GREENBERG, Jane. Metadata extraction and harvesting: a comparison of two automatic metadata generation applications. **Journal of Internet Cataloging**, [S. l.], v. 6, n. 4, p. 59-82, 2004. Disponível em: https://doi.org/10.1300/J141v06n04_05. Acesso em: 27 abr. 2024.

GRUBER, Thomas R. Towards principles for the design of ontologies used for knowledge sharing. **International Journal of Human-Computer Studies**, [S. l.], v. 43, n. 5-6, p. 907-928, nov. 1995. Disponível em: <https://doi.org/10.1006/ijhc.1995.1081>. Acesso em: 27 abr. 2024.

GS1. **Global Product Classification (GPC)**. 2024a. Disponível em: <https://www.gs1.org/standards/gpc>. Acesso em: 05 fev. 2024.

GS1. **GS1 Web Vocabulary**. 2024b. Disponível em: <https://www.gs1.org/voc/>. Acesso em: 05 fev. 2024.

GUARINO, Nicoa; OBERLE, Daniel; STAAB, Steffen. What is an Ontology? *In*: STAAB, Steffen; STUDER, Rudi. **Handbook on Ontologies**. 2. ed. Dordrecht: Springer, 2009. p. 1-17.

GUIMARÃES, José Augusto Chaves; TOGNOLI, Natália Bolfarini. Provenance as a domain analysis approach in archival knowledge organization. **Knowledge Organization**, [S. l.], v. 42, n. 8, p. 562-569, 2015. Disponível em: <https://repositorio.unesp.br/handle/11449/164830>. Acesso em: 11 nov. 2022.

GUISSONI, Leandro Angotti; CONSOLI, Matheus Alberto; RODRIGUES, Jonny Mateus. Is category management in small supermarkets worth the effort? **RAE**, São Paulo, v. 53, n. 6, p. 592-603, nov./dez. 2013. Disponível em: <https://doi.org/10.1590/S0034-75902013005000005>. Acesso em: 28 abr. 2024.

GUO, Jingzhi; SUN, Chengzheng. Collaborative product representation for emergent electronic marketplace. *In*: BLED ECONFERENCE, 12., 2003. **Proceedings** [...]. Bled: University of Maribor, 2003. Disponível em: <https://aisel.aisnet.org/bled2003/12>. Acesso em: 28 abr. 2024.

HEPP, Martin. **EClassOWL**: The web ontology for products and services. Neubiberg: UniBw München, 2023a. Disponível em: <http://www.heppnetz.de/projects/eclassowl/>. Acesso em: 09 jul. 2023.

HEPP, Martin. **The Product Types Ontology**: high-precision identifiers for product types based on Wikipedia. Neubiberg: UniBw München, 2023b. Disponível em: <http://www.productontology.org/>. Acesso em: 09 jul. 2023.

HIDER, Philip. **Information resource description**: creating and managing metadata. London: Facet Publishing, 2018.

HINDIANA. **Sobre nós**. São Paulo: Fundo Hindiana, 2022. Disponível em: <https://www.hindiana.com.br/>. Acesso em: 16 nov. 2022.

HJØRLAND, Birger. Does the traditional thesaurus have a place in modern information retrieval? **Knowledge Organization**, [S. l.], v. 43, n. 3, p. 145-159, 2016a. Disponível em: https://www.researchgate.net/publication/305264631_Does_the_Traditional_Thesaurus_Have_a_Place_in_Modern_Information_Retrieval. Acesso em: 19 dez. 2021.

HJØRLAND, Birger. Domain Analysis. **Knowledge Organization**, [S. l.], v. 44, n. 6, p. 436-464, 2017. Disponível em: https://www.isko.org/cyclo/domain_analysis. Acesso em: 11 nov. 2022.

HJØRLAND, Birger. Domain Analysis in Information Science: eleven approaches – traditional well as innovative. **Journal of Documentation**, [S. l.], v. 58, n. 4, p. 422-462, 2002. Disponível em: https://www.researchgate.net/publication/249366184_Domain_analysis_in_information_science_Eleven_approaches_-_Traditional_as_well_as_innovative. Acesso em: 11 nov. 2022.

HJØRLAND, Birger. Knowledge Organization. **Knowledge Organization**, [S. l.], v. 43, n. 6, p. 475-484, 2016b. Disponível em: https://www.ergon-verlag.de/isko_ko/downloads/ko_43_2016_6_j.pdf. Acesso em: 12 dez. 2021.

HJØRLAND, Birger. The classification of Psychology: a case study in the classification of a knowledge field. **Knowledge Organization**, [S. l.], v. 25, n. 4, p. 162-201, 1998. Disponível em: <https://repository.arizona.edu/handle/10150/105294>. Acesso em: 11 nov. 2022.

HJØRLAND, Birger; ALBRECHTSEN, Hanne. Toward a new horizon in Information Science: Domain-Analysis. **Journal of the American Society for Information Science**, [S. l.], v. 46, n. 6, p. 400-425, 1995. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-4571%28199507%2946%3A6%3C400%3A%3AAID-ASI2%3E3.0.CO%3B2-Y>. Acesso em: 11 nov. 2022.

HJØRLAND, Birger; BARROS, Thiago Henrique Bragato. Domain Analysis versus Facet Analysis. **Knowledge Organization**, [S. l.], v. 51, n. 1, p. 19-25, 2024. Disponível em: https://www.researchgate.net/publication/379227471_Domain_Analysis_Versus_Facet_Analysis. Acesso em: 19 ago. 2024.

HODGE, Gail. **Systems of knowledge organization for digital libraries**: beyond traditional authority files. Washington: The Digital Library Federation Council on Library Information Resources, 2000. Disponível em: <http://old.diglib.org/pubs/dlf090/dlf090.pdf>. Acesso em: 28 jul. 2023.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Comissão Nacional De Classificação. **Classificação Nacional de Atividades Econômicas**: estrutura. Rio de Janeiro: IBGE, 2023. Disponível em: https://cnae.ibge.gov.br/?option=com_cnae&view=estrutura&Itemid=6160&chave=&tipo=cnae&versao_classe=7.0.0&versao_subclasse=9.1.0&_gl=1*ws14us*_ga*MTg0MjAxMzYwMi4xNjg5NDYwNzY5*_ga_0VE4HSDTTT*MTY4OTczNTU0MS4zLjEuMTY4OTczNTg4MC4wLjAuMA. Acesso em: 19 jul. 2023.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Declaração dos Princípios Internacionais de Catalogação**. Haia: IFLA, 2016. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf. Acesso em: 09 jul. 2023.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional Requirements for Bibliographic Records**: final report. Haia: IFLA, 1997. Disponível em: <https://repository.ifla.org/handle/123456789/830>. Acesso em: 30 abr. 2024.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1**: information and documentation: thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneve: International Organization for Standardization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 704**: Terminology work: Principles and methods. Geneve: International Organization for Standardization, 2022.

JESUS, Samantha Augusta dos Santos de; MOREIRA, Walter. Contribuições teóricas da terminologia nos sistemas e processos de organização do conhecimento. *In*: EREBD: ENCONTRO REGIONAL DOS ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, GESTÃO E CIÊNCIA DA INFORMAÇÃO DAS REGIÕES SUDESTE, CENTRO-OESTE E SUL, 5., 2018, Belo Horizonte. **Anais** [...]. Belo Horizonte: UFMG, 2018. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/16907>. Acesso em: 30 abr. 2024.

KEJRIWAL, Mayank *et al.* Empirical best practices on using product-specific Schema.org. *In*: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 35., 2021, *online*. **Proceedings** [...]. [S. l.]: AAAI, 2021. Disponível em: <https://cdn.aaai.org/ojs/17816/17816-13-21310-1-2-20210518.pdf>. Acesso em: 30 abr. 2024.

KOTLER, Philip; KELLER, Kevin L. **Administração de marketing**. 14. ed. São Paulo: Pearson Education do Brasil, 2012.

KRIEGER, Maria da Graça. Terminologia e seus objetos de investigação. *In*: SIMPOSIO IBEROAMERICANO DE TERMINOLOGÍA, 10., 2008, Montevideo. **Actas** [...]. Montevideo: RITERM, 2008. p. 1-8

LAIPELT, Rita do Carmo Ferreira; MONTEIRO-KREBS, Luciana. **Termos sob a superfície**: elementos teóricos, metodológicos e terminológicos para a Representação do Conhecimento. Rio de Janeiro: Interciência, 2021.

LAMBE, Patrick. **Organising knowledge**: taxonomy, knowledge and organization effectiveness. Oxfordshire: Chandos Publishing, 2007.

LANGUAL. **LanguaL**: the International Framework for Food Description. Roskilde: LANGUAL, 2023. Disponível em: <https://www.langual.org/Default.asp>. Acesso em: 07 fev. 2024.

LARA, Marilda Lopes Ginez de. Novas relações entre Terminologia e Ciência da Informação na perspectiva de um conceito contemporâneo da informação.

DataGramZero, Rio de Janeiro, v. 7, n. 4, p. 1-9, ago. 2006. Disponível em: <https://www.brapci.inf.br/#/v/5938>. Acesso em: 30 abr. 2024.

LARA, Marilda Lopes Ginez de. Propostas de tipologias de KOS: uma análise das referências de formas dominantes de organização do conhecimento. **Encontros Bibli**, Florianópolis, v. 20, n. esp., p. 89-107, 2015. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20nesp1p89/28637>. Acesso em: 12 dez. 2021.

LARENTIS, Fabiano; SLONGO, Luiz Antonio; MILAN, Gabriel Sperandio. A relação entre o gerenciamento por categorias no varejo de auto-serviço e a prática do marketing de relacionamento. **REAd – Revista Eletrônica de Administração**, Porto Alegre, v. 12, n. 3, maio/jun. 2006. Disponível em: <https://seer.ufrgs.br/index.php/read/article/view/40559>. Acesso em: 30 abr. 2024.

LIMA, Gercina Ângela de. A análise facetada na modelagem conceitual de sistemas de hipertexto: uma revisão de literatura. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 7, n. 2, p. 189-196, jul./dez. 2002. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23431>. Acesso em: 30 abr. 2024.

LIMA, Gercina Ângela de. Representação do conhecimento e da informação no contexto da web. In: LIMA, Gercina Ângela de. **Organização e representação do conhecimento e da informação para web**: práticas metodológicas. Rio de Janeiro: Interciência, 2020. p. 1-20.

LIMA, José Leonardo Oliveira; ALVARES, Lilian. Organização e representação da informação e do conhecimento. In: ALVARES, Lilian (org.). **Organização da Informação e do Conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4, 2012. p. 21-48.

LISE, Fernanda (org.) *et al.* **Etapas da construção científica**: da curiosidade acadêmica à publicação dos resultados. Pelotas: Ed. UFPel, 2018. Disponível em: <http://guaiaca.ufpel.edu.br/handle/prefix/4171>. Acesso em: 30 abr. 2024.

LÓPEZ-HUERTAS, María J. Domain Analysis for interdisciplinary knowledge domains. **Knowledge Organization**, [S. l.], v. 42, n. 8, p. 570-580, 2015. Disponível em: https://www.researchgate.net/publication/298717384_Domain_Analysis_for_Interdisciplinary_Knowledge_Domains. Acesso em: 11 nov. 2022.

MACHADO, Dayane, Onaga Ferreira; ARAKAKI, Ana Carolina Simionato. Schema.org para catálogos digitais. **Encontros Bibli**, Florianópolis, v. 22, p. 01-21, 2022. Disponível em: <https://doi.org/10.5007/1518-2924.2022.e87046>. Acesso em: 30 abr. 2024.

MACHADO, Raildo de Sousa; ZAFALON, Zaira Regina. **Catálogo**: dos princípios e teorias ao RDA e IFLA LRM. João Pessoa: Editora UFPB, 2020.

MACULAN, Benildes Coura Moreira dos Santos; LIMA, Gercina Angela Borém de Oliveira. Taxonomia facetada navegacional: agregando valor às informações

disponibilizadas em bibliotecas digitais de teses e dissertações. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília, DF. **Anais** [...]. Brasília, DF: Enancib, 2011. p. 696-714.

MADSEN, Bodil Nistrup; THOMSEN, Hanne Erdman. Concept modeling vs. data modeling in practice. *In*: KOCKAERT, Hendrik J.; STEURS, Frieda. **Handbook of Terminology**. Amsterdam: John Benjamins Publishing Company, 2015. p. 250-275.

MAI, Jens-Erik. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, [S. l.], v. 41, p. 599–611, 2005. Disponível em: http://jenserikmai.info/Papers/2005_AnalysisInIndexing.pdf. Acesso em: 11 nov. 2022.

MARTINEZ, Ana *et al.* Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios web argentinos. **Ciência da Informação**, Brasília, DF, v. 33, n. 2, p. 106-111, ago. 2004. Disponível em: <https://doi.org/10.1590/S0100-19652004000200011>. Acesso em: 30 abr. 2024.

MATIAS-PEREIRA, José. **Manual de Metodologia da Pesquisa Científica**. Rio de Janeiro: Grupo GEN, 2016.

MATOS, Alexandre Isabel; BELFO, Fernando Paulo. Use of ontologies in Product Information Management: a proposal for a multinational engineering and technology company. **Procedia Computer Science**, [S. l.], v. 204, p. 599-609, 2022. Disponível em: <https://doi.org/10.1016/j.procs.2022.08.073>. Acesso em: 30 abr. 2024.

MAYERNIK, Matthew S. Metadata. **Knowledge Organization**, [S. l.], v. 47, n. 8, p. 696-713, 2020. Disponível em: <https://www.isko.org/cyclo/metadata>. Acesso em: 09 jul. 2023.

MENDONÇA, Fabrício Martins; ZAIDAN, Fernando Hadad. Ontologias para organização da informação em processos de transformação digital. **Em Questão**, Porto Alegre, v. 25, n. 1, p. 295-320, jan./abr. 2019. Disponível em: <https://doi.org/10.19132/1808-5245251.295-320>. Acesso em: 30 abr. 2024.

META. **Como adicionar uma categoria de produto do Google ou do Facebook para itens do catálogo**. 2024. Disponível em: <https://www.facebook.com/business/help/526764014610932?id=725943027795860>. Acesso em: 18 mar. 2024.

MEY, Eliane Serrão Alves. **Introdução à Catalogação**. Brasília, DF: Briquet de Lemos, 1995.

MICHEL, Maria H. **Metodologia e pesquisa científica em Ciências Sociais**. Rio de Janeiro: Grupo GEN, 2015.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **A framework of guidance for building good digital collections**. 3. ed. Baltimore: National Information Standards Organization, 2007. *E-book*. Disponível em: <http://www.niso.org/publications/rp/framework3.pdf>. Acesso em: 09 jul. 2023.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Baltimore: National Information Standards Organization, 2010.

NCM - Nomenclatura Comum do Mercosul. **Receita Federal**, Brasília, DF, 05 fev. 2021. Disponível em: <https://www.gov.br/receitafederal/pt-br/assuntos/aduana-e-comercio-exterior/classificacao-fiscal-de-mercadorias/ncm>. Acesso em: 19 jul. 2023.

NOY, Natalya F.; MCGUINNESS, Deborah L. Ontology development 101: a guide to creating your first ontology. **Knowledge Systems Laboratory**, [S. l.] n. 32, 2001. Disponível em: https://www.researchgate.net/publication/243772462_Ontology_Development_101_A_Guide_to_Creating_Your_First_Ontology. Acesso em: 14 fev. 2023.

PARENTE, Juracy. **Varejo no Brasil**: gestão e estratégia. São Paulo: Atlas, 2000.

PAVÃO, Caterina Groposo *et al.* Metadados e repositórios institucionais: uma relação indissociável para a qualidade da recuperação e visibilidade da informação. **PontodeAcesso**, Salvador, v. 9, n. 2, p. 103-116, dez. 2015. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/15163>. Acesso em: 30 abr. 2024.

PEREIRA, Djhonatan Baldivia; BARROS, João Paulo Pereira. **A importância do gerenciamento por categorias no setor de varejo supermercadista do Brasil**. 1. ed. Ourinhos: Faculdades Integradas de Ourinhos, 2017.

PINHO, Fabio Assis; VITAL, Luciane Paula. Classificação facetada aplicada em sistemas de organização e representação do conhecimento. *In*: LUCAS, Elaine Rosangela de Oliveira; CORRÊA, Elisa Cristina Delfini; EGGERT-STEINDEL, Gisela (org.). **As contribuições de Ranganathan para a Biblioteconomia**: reflexões e desafios. São Paulo: FEBAB, 2016. p. 96-107.

POMERANTZ, Jeffrey. **Metadata**. Cambridge: The MIT Press, 2015.

PONTES, Flavio Vieira; LIMA, Gercina Ângela Borém de Oliveira. A organização do conhecimento em ambientes digitais: aplicação da teoria da classificação facetada. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 4, p. 18-40, out./dez. 2012. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22806>. Acesso em: 30 abr. 2024.

PRIDE, William M.; FERREL, O C. **Fundamentos de Marketing**: conceitos e práticas. São Paulo: Cengage Learning Brasil, 2016.

PROTÉGÉ. **About**. Standford: Stanford University, 2023. Disponível em: <https://protege.stanford.edu/about.php>. Acesso em: 20 fev. 2023.

ROCHA, Marcos; MOLA, Jeferson. **Varejo**. São Paulo: Saraiva Educação, 2018.

SAAB, William George Lopes; GIMENEZ, Luiz Carlos Perez. Aspectos atuais do varejo de alimentos no mundo e no Brasil. **BNDES Setorial**, Rio de Janeiro, n. 11, p. 101-122, 2000. Disponível em: <http://web.bndes.gov.br/bib/jspui/handle/1408/10195>. Acesso em: 30 abr. 2024.

SALES, Rodrigo de; CAFÉ, Ligia. Tesouros e ontologias sob o olhar da Teoria Comunicativa da Terminologia. *In*: CONGRESSO ISKO SPAIN, 9., 2009, Valencia. **Anais [...]**. Valencia: ISKO SPAIN, 2009. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=2923286>. Acesso em: 30 abr. 2024.

SANTAROSA, Rosana. **A inatividade das lojas virtuais**: uma análise do comércio eletrônico à luz da inovação. 2016. 121 f. Dissertação (Mestrado em Administração de Empresas) – Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2016. Disponível em: <https://hdl.handle.net/10438/16223>. Acesso em: 30 abr. 2024.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SIMIONATO, Ana Carolina; ARAKAKI, Felipe Augusto. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, Londrina, v. 19, n. 1, p. 146-163, jan./abr. 2014. Disponível em: <https://doi.org/10.5433/1981-8920.2014v19n1p146>. Acesso em: 30 abr. 2024.

SA VAREJO. **Guia de sortimento**. São Paulo: SA Varejo, 2019.

SCHEMA BLOG. **GS1 Web vocabulary**: welcoming the first schema.org external extension. [S. l.], 22 fev. 2016. Disponível em: <http://blog.schema.org/2016/02/gs1-milestone-first-schemaorg-external.html>. Acesso em: 09 jul. 2023.

SCHEMA.ORG. **Data Model**. 2023a. Disponível em: <https://schema.org/docs/datamodel.html>. Acesso em: 09 jul. 2023.

SCHEMA.ORG. **Organization of Schemas**. 2023b. Disponível em: <https://schema.org/docs/schemas.html>. Acesso em: 09 jul. 2023.

SCHEMA.ORG. **Product**. 2023c. Disponível em: <https://schema.org/Product>. Acesso em: 09 jul. 2023.

SCHEMA.ORG. **Welcome to Schema.org**. 2023d. Disponível em: <https://schema.org/>. Acesso em: 09 jul. 2023.

SCHIESSL, Marcelo; SHINTAKU, Milton. Sistemas de Organização do Conhecimento. *In*: ALVARES, Lillian. **Organização da Informação e do Conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4, 2012. p. 49-118.

SERRA, Luís Henrique. A variação denominativa no discurso especializado do universo da cana de açúcar: a pertinência de dados orais na pesquisa terminológica. **Revista Digital dos Programas de Pós-Graduação do Departamento de Letras e Artes da UEFS**, Feira de Santana, v. 20, n. 1, p. 192-203, jan./abr. 2019. Disponível em: <https://doi.org/10.13102/cl.v20i1.4746>. Acesso em: 30 abr. 2024.

SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS. **Biscoitos caseiros / não industrializados**: estudos de mercado SEBRAE/ESPM 2008: relatório completo. Brasília, DF: SEBRAE, 2008.

SERVIÇO BRASILEIRO DE RESPOSTAS TÉCNICAS. **Resposta Técnica 6409**. Brasília, DF: SBRT, 2007.

SHAPER, Dudley. **Reason and the search for knowledge**: investigations in the philosophy of science. Boston: Reidel, 1984.

SIMÕES, Maria; LIMA, Gercina (coord.). **Do tratamento à organização da informação**: reflexões sobre concepções, perspectivas e tendências. Coimbra: Imprensa da Universidade de Coimbra, 2020.

SMIRAGLIA, Richard. **Domain analysis for knowledge organization**: tools for ontology extraction. Oxford: Chandos Publishing, 2015.

SMIRAGLIA, Richard. **Metadata**: a cataloger's primer. Washington: Outledge-Usa, 2005.

SOBRE O SELO Vegano. **Sociedade Vegetariana Brasileira**. São Paulo, 29 ago. 2023. Disponível em: <https://svb.org.br/selovegano/sobre-selo-vegano/>. Acesso em: 14 fev. 2024.

SOCIEDADE BRASILEIRA DE VAREJO E CONSUMO. **Ranking**: um estudo completo sobre o setor. 8. ed. São Paulo: Ed. Lamônica Conectada, 2022.

SOERTEL, Dagobert. **The representation of Knowledge Organization Structure (KOS) data**. a multiplicity of standards. Roanoke: JCDL, 2001. Disponível em: <http://www.dsoertel.com/cv/B75.pdf>. Acesso em: 28 jul. 2023.

SOUZA, Bruna; BREMGARTNER, Vitor. Evolução das modalidades B2B e B2C em e-business no Brasil. *In*: CONASUM, 11., 2016, Cuiabá. **Anais [...]**. Cuiabá: UFMT, 2016. Disponível em: <https://eventosacademicos.ufmt.br/index.php/CONASUM/IV-Conasum/paper/view/786>. Acesso em: 30 abr. 2024.

SOUZA, Renato Rocha; TUDHOPE, Douglas; ALMEIDA, Maurício Barcellos. Towards a Taxonomy of KOS: dimensions for classifying Knowledge Organization Systems. **Knowledge Organization**, [S. l.], v. 39, n. 3, p. 179-192, 2012. Disponível em: https://www.researchgate.net/publication/326071562_Towards_a_Taxonomy_of_KOS_Dimensions_for_Classifying_Knowledge_Organization_Systems. Acesso em: 30 abr. 2024.

STEIMER, Isadora dos Santos Garrido; LUZ, Charley dos Santos. Taxonomia para comércio eletrônico: diferentes perspectivas em *front e back end*. **Ciência da Informação em Revista**, Maceió, v. 2, n. 3, p. 3-14, 2015. Disponível em: <https://doi.org/10.28998/cirev.2015v2n3a>. Acesso em: 30 abr. 2024.

SUGIMOTO, Cassidy; WEINGART, Scott. The Kaleidoscope of disciplinarity. **Journal of Documentation**, [S. l.], v. 71, p. 775-94, 2015. Disponível em: https://www.researchgate.net/publication/281161918_The_kaleidoscope_of_disciplinarity. Acesso em: 11 nov. 2022.

TENNIS, Joseph T. Two axes of domains for Domain Analysis. **Knowledge Organization**, [S. l.], v. 30, n. 3-4, p. 191-195, 2003. Disponível em: https://www.researchgate.net/publication/244476999_Two_Axes_of_Domains_for_Domain_Analysis. Acesso em: 11 nov. 2022.

TENNIS, Joseph T. What does a domain analysis look like in form, function, and genre? **Brazilian Journal of Information Science**, Marília, v. 6, n. 1, p.3-14, 2012. Disponível em: <https://doi.org/10.36311/1981-1640.2012.v6n1.02.p3>. Acesso em: 11 nov. 2022.

TERCEDOR SÁNCHEZ, Maribel. Uma perspectiva situada de la variación denominativa. **Debate Terminológico**, n. 9, p. 81-88, fev. 2013. Disponível em: <https://seer.ufrgs.br/index.php/riterm/article/view/37173>. Acesso em: 30 abr. 2024.

TOP, Jan; TIMMER, Marielle; SIMSEK-SENEL, Görkem. **Food Data and Food Vocabularies**. Wageningen: Wageningen University & Research: 2020.

UNSPSC. **Welcome**. 2022. Disponível em: <https://www.unspsc.org/>. Acesso em: 19 jul. 2023.

VAJDA, Jonathan; MERRELL, Eric; SMITH, Barry. Toward an Ontology of Commercial Exchange. *In*: JOINT ONTOLOGY WORKSHOPS, 2019, Graz. **Proceedings** [...]. Graz: Medical University of Graz, 2019. Disponível em: <https://philarchive.org/rec/VAJTAO>. Acesso em: 30 abr. 2024.

VICKERY, B.C. **Faceted classification**: a guide to the construction and use of special schemes. Londres: Aslib, 1960.

VIEIRA, Gustavo Luiz Scatolini. Integração da cadeia de suprimentos no comércio eletrônico: estudo de casos múltiplos em pequenas e médias empresas no Brasil. **Revista Gest@o.Org**, Recife, v. 18, n. 1, p. 116-135, 2020. Disponível em: <https://doi.org/10.21714/1679-1827-2020.v18.n1.p116-135>. Acesso em: 16 nov. 2022.

VITAL, Luciane Paula. **Recomendações para construção de taxonomia em portais corporativos**. 2007. 113 f. Dissertação (Mestrado em Ciência da Informação) – Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em: <http://repositorio.ufsc.br/xmlui/handle/123456789/89791>. Acesso em: 30 abr. 2024.

VITAL, Luciane Paula; CAFÉ, Ligia Maria Arruda. Práticas de elaboração de taxonomias: análise e recomendações. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 8., 2007, Salvador. **Anais** [...]. Salvador: UFBA, 2007. Disponível em: <http://enancib.ppgci.ufba.br/artigos/GT2--141.pdf>. Acesso em: 30 abr. 2024.

W3C BRASIL. **Web Semântica**. 2023. Disponível em:
<https://www.w3c.br/Padroes/WebSemantica>. Acesso em: 15 fev. 2023.

WARBURTON, Kara. **The corporate terminologist**. Amsterdam: John Benjamins Publishing Company, 2021. (Terminology and Lexicography Research and Practice, 21).

ZENG, Marcia. Knowledge Organization Systems (KOS). **Knowledge Organization**, [S. l.], v. 35, n. 2-3, p. 160-182, 2008.

ZHONGHONG, W. *et al.* Potential and prospects of taxonomies for organization. **Knowledge Organization**, [S. l.], v. 33, n. 3, p. 160-169, 2006.

APÊNDICE A – Descrições dos produtos

DOI do arquivo no repositório digital Zenodo:

<https://doi.org/10.5281/zenodo.11081212>

APÊNDICE B – Termos extraídos das descrições dos produtos

DOI do arquivo no repositório digital Zenodo:

<https://doi.org/10.5281/zenodo.11081228>

APÊNDICE C – Definições dos termos extraídos

DOI do arquivo no repositório digital Zenodo:

<https://doi.org/10.5281/zenodo.11089740>

APÊNDICE D – Compilado de metadados de referência

DOI do arquivo no repositório digital Zenodo:

<https://doi.org/10.5281/zenodo.11081234>

APÊNDICE E – Modelo de representação completo em Excel

DOI do arquivo no repositório digital Zenodo:

<https://doi.org/10.5281/zenodo.11081251>