



## Metodologia para construção de taxonomias: uma revisão sistemática de literatura

### *Methodology for constructing taxonomies: a systematic literature review*

**Celia da Consolação Dias** 

Doutora em Ciência da Informação  
Universidade Federal de Minas Gerais, Brasil  
*celiadas@gmail.com*

**José Antonio Moreira González** 

Doutor em História  
Universidad Carlos III de Madrid, Espanha  
*jamore@bib.uc3m.es*

#### Resumo

Trata-se de um estudo sobre taxonomias aplicadas em diferentes contextos da sociedade com vistas a identificar os elementos da sua estrutura para avaliar a organização dos dados requeridos para a sua construção. O objetivo do artigo é identificar na literatura da Biblioteconomia e Ciência da Informação as metodologias para construção de taxonomias, suas etapas e aplicações, bem como avaliar e refletir sobre esse processo do ponto de vista dos componentes da estrutura. Os procedimentos metodológicos foram: 1) pesquisa bibliográfica com definição de protocolo estruturado de busca, com seleção dos termos no título, resumo e nas palavras-chave; 2) análise de conteúdo e definição de oito categorias estabelecidas a priori; 3) análise das informações do *corpus* da pesquisa com foco nos elementos da estrutura das taxonomias e da base teórica adotada. Destacam-se estudos com maior nível de detalhamento das etapas e ações, bem como das recomendações propostas pelos autores na execução de cada etapa, com ênfase nas atividades referentes a construção da estrutura das taxonomias. Salienta-se a importância do uso conjunto de procedimentos tecnológicos a partir do estabelecimento de procedimentos operacionais para organizar a construção dos componentes da estrutura da taxonomia.

**Palavras-chave:** taxonomias; metodologia de construção; revisão de literatura; identificação de tendências; sistemas de organização do conhecimento em empresas.

#### Abstract

*This is a study on taxonomies applied in different societal contexts, aiming to identify the elements of their structure to evaluate the organization of the data required for their construction. The objective of the article is to identify in the literature of Library and Information Science the methodologies for constructing taxonomies, their stages and applications, as well as to evaluate and reflect on this process from the perspective of structural components. The methodological procedures were: 1) bibliographic research with the definition of a structured search protocol, selecting terms in the title, abstract, and keywords; 2) content analysis and definition of eight categories established a priori; 3) analysis of the research corpus information focusing on the elements of the taxonomy structure and the theoretical basis adopted. Studies with a higher level of detail in the stages and actions, as well as the recommendations proposed by the authors in the execution of each stage, are highlighted, with an emphasis on activities related to the construction of the taxonomy structure. The importance of the combined use of*



doi: [10.28998/cirev.2024v11e18485](https://doi.org/10.28998/cirev.2024v11e18485)

Este artigo está licenciado sob uma [Licença Creative Commons 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

Submetido em: dd/mm/aaaa

Aceito em: dd/mm/aaaa

Publicado em: dd/mm/aaaa

*technological procedures is emphasized, starting from the establishment of operational procedures to organize the construction of the components of the taxonomy structure.*

**Keywords:** *taxonomies; construction methodology; literature review; trend identification; knowledge organization systems in companies.*

## 1 INTRODUÇÃO

Há muito tempo a sociedade tem sido exposta a uma enorme quantidade de publicação de dados de toda natureza que é produzida diariamente. Paralelo a esse cenário outras situações têm marcado esse momento, como a transformação digital experienciada pela sociedade e proporcionada pelas tecnologias de informação e comunicação, bem como o impacto causado pela presença da inteligência artificial e seu uso em atividades diversas têm demandado mudanças no cotidiano das organizações e das pessoas. Todas essas transformações mostram a necessidade de as organizações executarem estratégias para organizar seus acervos de dados e informação para se adequarem às exigências impostas pelos governos, pelos mercados e pela sociedade, decorrente do uso de tais recursos. Esse cenário compreende não apenas o planejamento de estratégias de negócios para efetivar a oferta de produtos, de serviços e de outras transações, mas também de informações que podem ser consumidas pela sociedade de maneira geral nas atividades diárias como cidadãos. Um recurso fundamental para realizar todas as atividades envolvidas nesse cenário é a informação e o conhecimento produzido pelas empresas, bem como aqueles requeridos na implementação dessas estratégias. Por outro lado, esses recursos demandam pela organização e classificação desse conjunto de informações para buscar, representar, conectar, padronizar, recuperar e descobrir informações.

Navegando por uma loja online pode-se perceber que há uma enorme oferta de produtos e serviços que atendem as diversas necessidades informativas daqueles que já têm em mente algo que estão buscando. Com a internet e as TIC novos recursos foram agregados aos dados sobre produtos e serviços das instituições. Entretanto, sabe-se que a oferta de informações sobre produtos e serviços requer pensar nos objetos e em estruturar os dados mediante vocabulários que os identifiquem, diferenciem, organizem e padronizem. Pode-se pensar, então que quando um cliente busca informação sobre determinado produto ou serviço quer interagir com os dados, tem interesse em saber nomes, detalhes técnicos sobre o produto e atributos como cor, preço, tamanhos, entre outros. Especificamente sobre esse quesito algumas lojas online têm tabelas auxiliares que apoiam o cliente na identificação do seu tamanho de determina pela de roupa, de calçado ou na correspondência entre os tamanhos usados em outros países. Sem esquecer a informação sobre os produtos mais vendidos, as avaliações dos clientes que compraram os produtos e até fotos de compradores usando um produto. Esse conjunto de dados revela a existência de um Sistema de Organização do Conhecimento (SOC) referencial formado por categorias, metadados, tabelas auxiliares e imagens dos produtos associados aos textos, com vistas a apoiar os clientes a navegar ou buscar informação de interesse. É um cenário onde o SOC adequado se estrutura como uma taxonomia de produtos. As taxonomias são estruturas classificatórias usadas para organização do conhecimento e de objetos de conteúdo, comumente usadas no ambiente digital e conforme apontado por Gomes (2014) como um elemento da arquitetura da informação e das ontologias.

Saracevic (1996) afirma que a origem da Ciência da Informação ocorreu no bojo da revolução científica e técnica que se seguiu à Segunda Guerra Mundial, em um cenário marcado

pela explosão informacional. Em seu artigo o autor menciona a visão de Bush em 1945, acerca dum problema de então que continua vigente no atual momento. Trata-se da necessidade de tornar acessível a massiva quantidade de informação de um acervo de conhecimento crescente. Esse aspecto é apontado por Vital (2007) e reforçado no estudo de Andriani (2017) ao reportar-se à função das taxonomias na oportunidade de organizar informações em ambientes digitais, destinadas a públicos distintos. Vital (2007) denota que diversos autores relacionam o aparecimento e o uso do termo em ambientes digitais ao surgimento de formas automatizadas de criação de informação. Andriani (2017) chama atenção que essa adoção decorre sobretudo pela quantidade de informação catalogada de forma digital ter aumentado exponencialmente, sendo necessários instrumentos cada vez mais inteligentes de filtragem para recuperá-las de forma eficaz.

Pode-se notar que tal problemática continua parecida com o cenário apontado por Saracevic (1996). Desde então, alcançou-se um nível diferente de oferta de TIC e a sociedade deparou-se com outras questões referentes a forma de se comunicar digitalmente em uma grande rede de informação. Nesse sentido, mudanças ocorreram notadamente marcadas pela abertura da *Word Wide Web* no final do século 20 e pela contínua transformação da internet, inicialmente caracterizada por páginas estáticas e com pouca interatividade. Mas, com o passar do tempo a sociedade passou a contar com “sites dinâmicos, plataformas de mídia social e uma mudança profunda em direção ao conteúdo gerado pelo usuário”, conforme apontado por Souza (2023). A autora relata que até os dias atuais “a Web 3.0, definida pela descentralização, tecnologia *blockchain*, o que garante transparência e segurança em transações online, compartilhamento de dados e a integração da inteligência artificial (IA) em nossas experiências online”, conforme apontado por Souza (2023). Esse cenário não só motivou as pessoas a criarem conteúdos, extrair dados dos objetos de conteúdos, disseminá-los pelas redes sociais, mas também usá-los para a tomada de decisão, para a geração de novos conhecimentos e de negócios, e a usar a potência da rede de computadores para a comunicação com distintos objetivos desde transações comerciais, como plataformas educacionais e inúmeras atividades realizadas pela sociedade, cenário fortemente desenvolvido sobretudo durante a pandemia da COVID de 2019.

Mas, mudanças adicionais ocorreram e impactaram as atividades associadas ao fazer dos profissionais responsáveis pela seleção, organização, disponibilização e recuperação de informação. Tais mudanças sucederam nos formatos e na representação dos objetos, em especial respeito à necessidade de alteração dos formatos de catalogação e dos padrões de metadados aplicáveis aos objetos de conteúdo. Do ponto de vista da construção dos instrumentos de organização do conhecimento essas mudanças contribuíram para uma evolução dos vocabulários a partir de funções como publicação de dados na web, conexão de objetos, interoperabilidade entre os vocabulários e qualquer outra possibilidade de integrar e compartilhar recursos na web (ISO 25964-2, 2013).

A temática é de interesse para a Ciência da Informação pela sua importância como SOC e porque as taxonomias representam um vocabulário padronizado e estruturado aplicável não só nos diversos campos de conhecimento, mas em especial nas atividades das organizações, de maneira geral. Sua prática é, portanto, multidisciplinar. O estudo faz parte de um projeto de pesquisa de pós-doutorado mais amplo que investiga a estrutura das taxonomias.

Nesse sentido, a seguinte questão norteou a realização desta pesquisa: Quais são as etapas seguidas para a construção de taxonomias e quais as bases teóricas que sustentam tal criação? Em particular, este estudo tem como objetivo: identificar na literatura da área de

biblioteconomia e Ciência da Informação as metodologias para construção de taxonomias, suas etapas e aplicações, bem como refletir sobre esse processo do ponto de vista dos componentes da sua estrutura.

O texto se expõe da seguinte forma: além da introdução, na seção 2 trata, brevemente, da construção de taxonomia. Assim como se oferecem adotada. A seção 3 apresenta informações sobre a metodologia de pesquisa. A seção 4 apresenta a análise de conteúdo dos dados e discussão dos resultados a partir da classificação e descrição de cada metodologia do corpus para a construção de taxonomias em oito categorias, com uma análise dos resultados com foco nos elementos da estrutura das taxonomias e da base teórica para a realização dos respectivos estudos da metodologia de construção de taxonomias. Por fim, a seção 5 traz as considerações finais e na sequência apresenta-se as referências.

## **2 REFERENCIAL TEÓRICO**

Observou-se que a literatura da Biblioteconomia e Ciência da Informação registra um amplo aporte teórico conceitual metodológico para a construção dos tesouros. Em comparação com os tesouros, essa situação não se repete para a construção das taxonomias, como observado por Suárez Sánchez; Hernández (2020). Os autores concordam sobre a ausência de normas técnicas, entretanto em algumas situações eles fazem referência às recomendações das normas de construção de tesouro para indicar que determinado procedimento deve ser adotado para uma situação similar a que ocorre para a construção das taxonomias.

Sobre as práticas de elaboração de taxonomias Vital & Café (2007), Aganete (2010) mencionam que “as metodologias atualmente disponíveis não apresentam orientações suficientemente detalhadas, o que deixa lacunas para entendimentos indevidos na sua construção e aplicação”. Nesse sentido, o estudo de Vital & Café (2007) focou a análise das práticas de elaboração sob a ótica da funcionalidade da categorização nos aspectos de: 1) Construção da taxonomia (representação do conhecimento e 2) Aplicação da taxonomia. Segundo as autoras o conceito de categorização envolve tanto aspectos da construção de taxonomia (representação do conhecimento) como da aplicação da taxonomia na representação da informação e organização da informação recuperada.

Ainda sobre a construção de taxonomias, observou-se que a literatura do campo de conhecimento, como já foi mencionado, contém relatos de experiência resultantes da prática e de estudos de investigação acerca de metodologias para a construção de taxonomia. A análise de tais estudos é o foco desse artigo. A seguir serão apresentados os procedimentos metodológicos para atingir o objetivo proposto.

## **3 PROCEDIMENTOS METODOLÓGICOS**

A pesquisa é de natureza bibliográfica descritiva e exploratória, com uma abordagem qualitativa acerca das metodologias de construção de taxonomias. Foi realizada com vistas a identificar na literatura da Ciência da Informação as publicações sobre as metodologias para construção de taxonomias para compreender quais as etapas do processo são indicadas nos estudos investigados. Os seguintes procedimentos metodológicos foram aplicados para atingir os objetivos desta investigação.

Quadro 1 - Protocolo de Revisão de Literatura

1) Delimitação do Assunto				
<b>Especificação do objeto da revisão</b>				
A busca será realizada em bases nacionais e internacionais das áreas de Biblioteconomia e Ciência da Informação, nos idiomas inglês, português e espanhol.				
<b>Redação da questão de revisão</b>				
Quais são as etapas seguidas para a construção de taxonomias e quais as bases teóricas que sustentam tal criação?				
2) Pesquisa Bibliográfica				
Fontes de Pesquisa	Biblioteconomia e Ciência da Informação	BRAPCI / CLASE: Citas Latinoamericanas en Ciencias Sociales y Humanidades / CrossRef / Dimensions / ENANCIB / E-prints in Library and Information Science (E-LIS) / Google Acadêmico / ISKO BRASIL / OAlster / OASIS / Perspectivas em Ciência da Informação / Repositório Institucional Digital do IBICT (RID1) / Repositório Institucional UFMG / Sage Journals / Scielo / Science Direct / Scopus / Web of Science / Springer Link / Portal Capes (Bases que indexam normas técnicas)		
Formulação da Expressão de Busca				
<b>Instrumento Terminológico</b>		Tesouro Brasileiro de Ciência da Informação		
<b>Palavras-chave</b>	Taxonomia	Construção de Taxonomias	Avaliação de Taxonomias	Sistema de Organização do Conhecimento
Controle Terminológico				
<b>Termo Livre nº 1</b>	<u>Português</u>	<u>Inglês</u>	<u>Espanhol</u>	
	Taxonomia	Taxonomy	Taxonomía	
<b>Descritores</b>	Taxonomias	Taxonomies	-	
<b>Termo Livre nº 2</b>	Construção de Taxonomias	Construction of Taxonomies	Construcción de taxonomías	
<b>Descritores</b>	Elaboração de linguagens documentárias	Index language construction	Construcción de vocabularios controlados	
Lógica das Expressões de Busca		Estrutura das Expressões de Busca		
<i>String 1</i>	2ª Expressão <i>Strings 1 + 2</i>	(("Taxonomia" OR "Taxonomy" OR "Taxonomía") AND ("Construção de taxonomia" OR "Construction of Taxonomies" OR "Construcción de taxonomías" OR "elaboração de linguagens documentárias" OR "index language construction" OR "construcción de vocabularios controlados"))		
<i>String 2</i>	2ª Expressão <i>String 2</i>	("Construção de taxonomia" OR "Construction of Taxonomies" OR "Construcción de taxonomías" OR "elaboração de linguagens documentárias" OR "index language construction" OR "construcción de vocabularios controlados")		

Fonte: elaboração própria (2024)

**Etapa 1:** Pesquisa exploratória e revisão sistemática de literatura: Registra-se, a seguir, o protocolo de revisão de literatura elaborado para o desenvolvimento da revisão de literatura. O protocolo contém a indicação dos componentes da pesquisa sem definição de recorte temporal e em três idiomas: português, inglês e espanhol. O quadro 1 contém os elementos do protocolo que nortearam a construção da estratégia de busca para a pesquisa bibliográfica: as fontes de informação selecionadas, o uso do Tesouro Brasileiro em Ciência da Informação (TBCI) para padronizar as palavras-chave e os termos usados para a busca, a definição das *strings* de busca com vistas a dar uma ampla cobertura temática. Para subsidiar a seleção das publicações recuperadas foram estabelecidos critérios de inclusão e exclusão dos itens recuperados. Os critérios de inclusão que foram considerados, são: textos completos e em acesso aberto; das áreas do conhecimento: de Biblioteconomia e Ciência da Informação. A seleção dos títulos de interesse foi realizada verificando a presença dos termos no título, resumo e nas palavras-chave.

**Etapa 2:** Análise de conteúdo, seguindo as recomendações de Bardin (2015) a partir da leitura de cada publicação do *corpus* da pesquisa, classificando os dados de cada metodologia para a construção de taxonomias e registrando os achados nas seguintes categorias de análise: 1. Publicação; 2. Objetivo do estudo; 3. Ambiente de realização/ foco da pesquisa; 4. Metodologia de construção de taxonomias; 5. Abordagem da estrutura; 6. Equipe; 7. Base teórica; e 8. Resultados da pesquisa. Etapa 3: Análise das informações coletadas na etapa anterior com foco nos elementos da estrutura das taxonomias e da base teórica adotada pelos autores para a realização dos respectivos estudos das metodologias de construção de taxonomias.

#### 4 RESULTADOS: ANÁLISES E DISCUSSÕES

Os estudos de Martínez, et. al (2004); Andriani (2017); Suárez Sánchez; y López Hernández (2020); Centelles (2005); Arudo; Centelles (2005); Hurtado; Losada (2011); Viana; Aganette; Maculan (2021); Aganette (2010) e Moreiro González (2005), conforme indicado no quadro 2, foram selecionados para compor o *corpus* de análise. Uma visão geral sobre as publicações revela que a temática do ponto de vista da abordagem deste estudo é pouco explorada na literatura da Biblioteconomia e da Ciência da Informação. Nota-se que, embora o levantamento tenha sido feito sem um recorte temporal definido, há um espaçamento cronológico entre as publicações. Por outro lado, pode-se inferir, também que a necessidade de publicar acerca da temática se deve a forte presença na internet das instituições de forma geral para aumentar a visibilidade da oferta de produtos e serviços, para realizar transações, desenvolver relacionamento entre públicos diversos nas redes sociais, para o compartilhamento de informações, entre outras ações proporcionadas pelas plataformas digitais. Ressalta-se que tais observações estão limitadas às publicações identificadas nas bases de dados consultadas e que apresentaram a seguinte configuração quantitativa, como apresentado no quadro 2.

Destaca-se que o primeiro estudo do *corpus* data de 2004, na sequência há dois artigos publicados em 2005, sendo que dois dos autores se repetem em duas publicações, uma dissertação em 2010, um artigo e um livro de 2011, uma dissertação em 2017, um artigo em 2020 e um em 2021, totalizando nove publicações. Observou-se que o levantamento bibliográfico resultou em forte presença dos pesquisadores espanhóis, autores de 67% das publicações cinco artigos e um livro. Ressalta-se que um dos pesquisadores é espanhol e autor individual de um livro publicado em português. São 37,5% das publicações publicadas no idioma português, formado por duas dissertações uma em 2010 e uma em 2017 e um artigo em 2021.

Qualitativamente os resultados foram apresentados usando oito categorias de análise para cada uma das publicações avaliadas nesta pesquisa: 1. Publicação: Indica o tipo de publicação; 2. Objetivo do estudo: Indica o objetivo principal do estudo; 3. Ambiente de realização/foco da pesquisa: Compreende informação sobre o campo do conhecimento ou de aplicação da taxonomia; 4. Metodologia de construção de taxonomia: Indica o passos a passo

com as orientações do procedimento metodológico usado; 5. Abordagem da estrutura: indica elementos usados da estrutura usados na construção da taxonomia; 6. Base teórica: compreende a indicação dos princípios teóricos para a criação da taxonomia proposta. 7. Equipe: registra se há menção à equipe responsável pelo desenvolvimento da taxonomia e 8. Resultados da pesquisa: contém os resultados diretamente relacionados com a temática investigada. Os resultados aqui apresentados estão focados nos elementos da estrutura das taxonomias e na base teórica adotada pelos autores para a realização dos respectivos estudos.

Quadro 2 – Publicações do *corpus* analisado

Tipo de Publicação	Título	Idioma	Ano
1) Artigo de periódico	Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios Web argentinos.	Espanhol	2004
2) Dissertação	Um método para a construção de taxonomias utilizando a DBPEDIA.	Português	2017
3) Artigo de periódico	Taxonomías digitales: fundamentos teóricos y metodología de construcción.	Espanhol	2020
4) Artigo de periódico	Metodología para el diseño de taxonomías corporativas.	Espanhol	2005
5) Artigo de periódico	Taxonomías para la categorización y la organización de la información en sitios web [en línea].	Espanhol	2005
6) Artigo de periódico	Método para la construcción de una taxonomía: estructura base para riesgos en outsourcing de software	Espanhol	2011
7) Dissertação	Taxonomias corporativas [manuscrito]: um estudo sobre definições e etapas de construção fundamentado na literatura publicada	Português	2010
8) Artigo de periódico	Metodologia para construção de taxonomias corporativas facetadas.	Português	2021
9) Capítulo de Livro	Taxonomia	Português	2011

Fonte: elaboração própria (2024)

O estudo de Martínez et. al. (2004) aborda o uso das categorias fundamentais do CRG para a construção de taxonomias. A ideia do estudo é analisar a potencialidade das facetas e com isso identificar se as categorias do CRG podem ser recomendadas para o desenho de sítios organizacionais na web. A pesquisa tem como objetivo identificar e analisar as categorias fundamentais que subjazem as diversas taxonomias de sítios na Web com vistas a determinar se coincidem ou não com as categorias fundamentais do CRG. O ambiente de realização da pesquisa compreendeu a análise de 25 empresas oriundas de diferentes atividades de atuação, tais como: três bodegas de vinho, três indústrias alimentícias, seis clubes de futebol, cinco associações profissionais, três faculdades universitárias e cinco agencias de governo. Como procedimento metodológico para construir a taxonomia foram definidas quatro etapas, a primeira é a análise das taxonomias selecionadas, com identificação das classes de primeiro e de segundo nível; como segunda etapa se realiza a classificação dessas classes conforme as categorias do CRG; após esse procedimento foi realizado o reagrupamento das classes em suas respectivas categorias gerando seis grupos de sítio distribuídos como (Martínez et. al., 2004): a. Bodegas de vino: tres sitios *Web*; b. Industrias alimenticias: tres sitios *Web*; c. Clubes de fútbol: seis sitios *Web*; d. Asociaciones profesionales: cinco sitios *Web*; e. Facultades universitarias: tres

sítios *Web*; f. Agencias de gobierno: cinco sítios *web*. E, na quarta etapa foi realizada uma análise dos seis grupos a fim de identificar se há a incidência das categorias fundamentais do CRG nos diferentes sítios analisados.

Como resultado destaca-se que, como o próprio CRG bem estabelece, nem todas as categorias são úteis ou necessárias para todas as organizações. Os autores ressaltam que as entidades da natureza ou entidades e organismos vivos dificilmente têm relacionamento com associações, agências profissionais ou agências do governo. Como conclusões, eles afirmam que as categorias fundamentais do CRG são adequadas para analisar as informações e construir taxonomias corporativas, contribuindo para otimizar os sítios *Web* correspondentes. Em segundo lugar, a ocorrência das indústrias alimentícias, faculdades e agencias de governo incluem ambas (100% das facetas), enquanto o sítio dos clubes de futebol e das associações profissionais só apresentaram uma das facetas. Esses resultados confirmam a aplicação das taxonomias como esquemas de classificação para organizar e recuperar informações de diferentes campos do conhecimento ou de atividade, assim como comprovam o potencial de uso das categorias do CRG como categorias. Embora o estudo apresente tabelas com os dados da análise a partir das categorias do CRG, observou-se que a metodologia proposta identificou as etapas de forma sintética e não apresenta os detalhes da execução de cada etapa. A teoria da classificação facetada de Ranganathan e os estudos do Grupo de Pesquisa em Classificação (CRG) fundamentaram os procedimentos metodológicos e as análises de Martínez, et. al. (2004).

A dissertação de Andriani (2017), mestrado em Engenharia do Conhecimento intitulada “Um método para a construção de taxonomias utilizando a DBPEDIA” tem como objetivo desenvolver um método para a extração de taxonomias hierarquizadas a partir de documentos, segundo a base de conhecimento DBPEDIA. O escopo da pesquisa está focado nas relações hierárquicas dos termos da taxonomia: o desenvolvimento de um método de construção de taxonomias que trate da identificação dos termos e da elaboração da hierarquia taxonômica. O ambiente de realização da pesquisa foi a Plataforma Lattes do CNPq, a partir do uso de uma amostra de currículos previamente definida. Observou-se que três dos quatro objetivos específicos identificados na publicação têm relação direta com o processo empírico de construção de taxonomias. A investigação teve como objetivo geral “propor um método automatizado para a construção de taxonomias, a partir do referencial teórico levantado na revisão e na análise; implementar em formato de *software* uma versão do método que demonstre a sua viabilidade e avaliar a funcionalidade do método, comparando os resultados obtidos com uma taxonomia de referência”(Andriani, 2017). O autor trabalhou um método para a geração de taxonomias a partir de repositórios de informações textuais com o apoio de bases de conhecimento, que fornecem as relações hierárquicas para verificação das relações taxonômicas entre os termos. O método proposto pelo autor compreende: 1.Extração de termos a serem utilizados pela taxonomia. A partir desses currículos é realizado um processo de reconhecimento de entidades, que extrai os principais termos mencionados. 2. Verificação das relações entre os termos extraídos anteriormente: Cada par de termos é validado e é encaminhado para a Dbpedia. Caso exista uma relação hierárquica válida entre as entidades (de acordo com os parâmetros definidos ao utilizar o método), ela é adicionada à árvore taxonômica. 3.Geração da árvore taxonômica: as relações são organizadas em formato de árvore, respeitando a hierarquia entre os termos. Ressalta-se que Andriani (2017) elaborou um método automatizado para a geração de taxonomias a partir de repositórios de informações textuais com duas variações de algoritmo-base, que tratam para testar a validade de relações taxonômicas: 1) Método baseado em coocorrências dentro do *corpus*: nesse caso a



ocorrência de um par de termos dentro do *corpus* basta para que se teste a existência de uma relação hierárquica entre elas na DBpedia; e 2) Método baseado em coocorrências dentro do currículo: nesta variação, a coocorrência dos termos dentro de cada currículo é utilizada como critério de validação para identificar possíveis relações hierárquicas. O método criado foi avaliado por meio da comparação dos resultados obtidos do experimento com uma taxonomia de referência, no caso o tesauro AGROVOC.

Sobre a equipe participante na construção das taxonomias se deve considerar sua composição por diversos perfis profissionais com competências distintas, visto que uma taxonomia que atende a um domínio completo depende da conjunção cognitiva multidisciplinar. Somente com a junção de competências é possível selecionar os conceitos de forma que estes reúnam todas as informações relevantes para a organização (Andriani, 2017, p. 45).

Em relação aos resultados alcançados com a pesquisa se reconhece que o objetivo geral da pesquisa foi atingido. Pois, partindo dos experimentos realizados, foi possível verificar que o método desenvolvido permite automatizar parcial ou plenamente a construção de taxonomias. Outro resultado relevante tocante ao método desenvolvido é que sua aplicação “também permite a geração de taxonomias em outros idiomas, bastando para isso reunir os modelos de reconhecimento de entidades e as árvores categóricas da Dbpedia específicos desses idiomas (Andriani, 2017, p. 115, 117). A fundamentação teórica usada da pesquisa se apoia na Ciência da Informação e compreende temáticas da área de organização e representação do conhecimento, taxonomias: características, tipos, sua estrutura e abordagens para a construção de taxonomias, normas internacionais de construção de tesouros e bases de conhecimento.

O artigo de Suárez Sánchez e López Hernández (2020) intitulado “Taxonomías digitales: fundamentos teóricos y metodología de construcción” tem como objetivo construir uma taxonomia digital do Sistema Solar que sirva como ferramenta de acesso temático, hierárquico, visual a um conjunto de recursos digitais (imagens, vídeos e textos) em um sítio web. O estudo pressupõe que as taxonomias digitais podem fazer esta organização sob nós conceituais-categoriais, assim como reforçado na definição de Hodge (2000, p. 6) que marca as taxonomias como um “sistema para a organização do conhecimento com características de notação conceitual e estruturação hierárquica que possibilitam representar temas amplos na rede”. Neste artigo a construção da taxonomia proposta por Suárez Sánchez e López Hernández (2020) é resultado da combinação de etapas e elementos das metodologias de Centelles (2005) e Hedden (2010), <sup>1</sup>como indicado a seguir: 1) Miquel Centelles (2005, p. 11) já constata que “as aplicações das taxonomias no contexto da web podem ser diversas”. Pelo que recomenda seguir quatro etapas em sua construção: (1) determinação da realidade; (2) extração do conjunto de termos ou das categorias; (3) padronização ou controle terminológico dos termos ou categorias; (4) estabelecimento do esquema e estrutura organizacional dos termos ou categorias. 2) Por sua vez, Heather Hedden (2010, p. XII-XIII) afirma que não existem cursos específicos para construir taxonomias, mas considera que é um assunto de interesse para múltiplos perfis profissionais: bibliotecários, arquitetos da informação, indexadores e publicitários, assim como gestores de informação e do conhecimento. Todos eles podem integrar equipes interdisciplinares que desenhem e elaborem taxonomias. Em relação ao processo de construção Hedden considera três grandes fases: (1) estabelecimento dos termos; (2) criação da estrutura categorial; e (3) estruturação da taxonomia mediante alguma ferramenta digital. As etapas da proposta de Suárez Sánchez e López Hernández (2020) compreendem: 1. Determinação da realidade: é a determina-

---

<sup>1</sup> Destaca-se que o livro da referida autora não foi selecionado para esta pesquisa por não estar disponível em acesso aberto, mas, sem dúvida, é uma obra de referência obrigatória sobre a construção de taxonomias, especialmente as comerciais. Heather Hedden projeta, cria e edita taxonomias, tesouros, metadados e ontologias para indexação e marcação de conteúdo para apoiar recuperação, pesquisa e localização de conteúdo.

ção de entidades, áreas de conhecimento, setores industriais, etc. que será representado pela taxonomia (Centelles, 2005). 2. Estabelecimento de termos: envolve obter informações sobre quais podem ser incluídos e, ainda, avaliar para decidir quais devem ser incluídos (Hedden, 2010). 3. Normalização de termos: consiste em duas tarefas: determinar os termos preferidos versus aqueles que não o são e dar uma forma correta e consistente a todos os termos da taxonomia (Centelles, 2005). 4. Criação de categorias: é a estruturação da taxonomia, considerando relações de equivalência, hierarquia, associação e semânticidade (Hedden, 2010). 5. Estruturação em um programa: um vasto grupo de taxonomias ainda é criado e mantido no Microsoft Excel, mas quando a taxonomia se torna muito complexa ou grande, como quando contém várias centenas de termos, é hora de movê-la para um software de taxonomia (Hedden, 2010). Suárez Sánchez; López Hernández, (2020) mencionam claramente a importância de uma equipe formada por bibliotecários, publicitários e arquitetos da informação, indexadores e gestores de informação e do conhecimento.

Os resultados se apresentam em cada uma das fases de construção da taxonomia. A implementação de cada etapa proposta requer conhecer as possíveis soluções e uma tomada de decisão acerca do procedimento mais adequado. Como por exemplo, na etapa de extração de termos da taxonomia sobre o Sistema solar os autores recomendam o uso de terminologias para o mapeamento dos termos que podem ser encontrados em Listas de cabeçalhos de assuntos, dicionários especializados na área de conhecimento de astronomia e cosmologia entre outros. Como resultado da aplicação da metodologia a taxonomia consiste em 55 elementos estruturados em oito nós categóricos que se granulam em quatro níveis de classe. A metodologia aplicada é considerada “altamente eficaz na execução de todo o processo de construção do instrumento, desde a determinação da realidade até a sua estruturação em programa” (Suárez; López, 2020). Os autores consideraram a metodologia proposta como excelentes para elaborar taxonomias digitais. De jeito que consideram ter atingido o objetivo estabelecido. A fundamentação teórica da proposta de construção da taxonomia está focada nas etapas e nos elementos propostos por Centelles (2005) e por (Hedden, 2010), em recomendações das norma *ISO para construção de tesouros no aspecto da interoperabilidade entre vocabulários e recuperação da informação abordados nas ISO25964*. Para a criação de categorias (Glushko, 2013 p. 244). Por fim os autores também se fundamentaram em autores da Ciência da Informação quando trataram sobre as Garantias de validação do instrumento e a referência para a lista de opções dos programas para construção da taxonomia do sistema solar.

A publicação de Arudo y Centelles (2005) situa o foco da investigação no desenho e elaboração de taxonomias corporativas afirma que “o que se pretende com o artigo é aportar alguma luz ao significado do termo taxonomia corporativa, assim como realizar uma breve aproximação a metodologia de construção desta ferramenta como base para organizar conteúdos digitais em ambiente corporativos”.

De início, confirmam a importância de dispor de uma equipe de profissionais com diferentes competências, informação, conhecimento e experiência que inclua bibliotecários e cientistas da informação com conhecimento específico das atividades e do funcionamento de uma organização concreta. E que agrupe, também, pessoal de cargos diretivos junto a representantes dos diversos departamentos da instituição, de tecnologia da informação e de desenho gráfico. Ainda em relação à essa equipe, formada por pessoas com diferentes visões e competências, o mais importante, segundo os autores, é conseguir que os envolvidos compreendam a importância da realização da tarefa e para isso se necessita compromisso, responsabilidade e disposição em participar e colaborar.

Uma boa elaboração de taxonomia, segundo Arudo y Centelles (2005), requer um detalhado desenho a cumprir rigorosamente, que o artigo determina em cinco etapas: 1) Planejamento estratégico que define e caracteriza a taxonomia a ser criada a partir da análise do contexto, público usuário e conteúdo; 2) Construção da estrutura: etapa que se recomenda considerar os fundamentos das normas de construção dos tesouros no que se refere ao controle de vocabulário e às relações entre conceitos. Sobre os elementos da estrutura pode se desatacar: 2.1) Uso da classificação facetada; 2.2) Identificação e estruturação das relações hierárquicas que vão conectar todas as categorias de uma faceta e representam com os hipônimos e hipônimos suas partes e instâncias; sem esquecer das relações associativas que conectam categorias de diferentes classes da taxonomia. Se destaca a recomendação para que o público participe desta etapa; 3) Categorização dos conteúdos: compreende o uso de categorização manual, automática ou semiautomática; 4) Aplicação e apresentação, ciclo de avaliação, melhoria contínua e 5) difusão. Em relação à equipe Arudo; Centelles (2005) afirmam que indivíduos com diferentes competências, conhecimentos e pontos de vista: especialistas na área de biblioteconomia e documentação, conhecedores do funcionamento da organização, cientistas da computação, designers gráficos, gestores de cargos, representantes dos vários departamentos ou unidades da organização, são fundamentais para a construção de taxonomias.

Como resultado se percebe que a metodologia proposta é abrangente, completa e oferece detalhes acerca dos procedimentos de cada etapa específica. Ressalta-se ainda a posição dos autores sobre as técnicas tradicionais de construção de tesouros como base teórica, bem como as recomendações para à construção de taxonomias, junto a outros pontos específicos aplicáveis às taxonomias corporativas. Também é ponto de destaque na pesquisa e que não foi observado em outros estudos do *corpus*, a recomendação de registrar os critérios estabelecidos durante o processo em um manual de diretrizes, pois facilitará a manutenção e os possíveis novos ajustes devidos a constante avaliação da taxonomia.

O artigo de Centelles (2005) não contém um objetivo claro, mas analisando a ideia central se deduz que se trata de propor uma definição do termo taxonomia, visitar a literatura sobre os processos de construção de taxonomias, sua aplicação na categorização de recursos e no desenvolvimento de sistemas de busca de informação para websites. O artigo parte da definição que faz a norma ANSI/NISO Z39.19 (2005) do termo Taxonomia como vocabulário controlado: um conjunto organizado de palavras ou frases usadas para organizar informações e destinadas principalmente à navegação mais do que à busca. Em consequência, por sua finalidade e seu ambiente de aplicação, a taxonomia se caracteriza enquanto vocabulário que “prioriza a exploração”. E que, portanto, encontra na web o seu ambiente de aplicação, com forte presença no “ambiente digital” (Centelles, 2005). Em relação aos outros tipos de vocabulários a norma aponta que as taxonomias incorporam relações de hierarquia e de equivalência como componentes semânticos da sua estrutura.

Já no tocante à metodologia para construir taxonomias o artigo chama a atenção para elaborar um planejamento estratégico que determine quais características a taxonomia deverá apresentar a partir da análise do contexto “que identificará as prioridades da corporação na organização e apresentação das informações no site”, do público alvo “que identificará as necessidades e comportamentos de pesquisa e uso de informações por diferentes segmentos de usuários e do conteúdo a representar (Centelles, 2005, p. 4). O autor propõe quatro etapas no processo de construção: 1) Delimitação da realidade (entidade, área do conhecimento, setor industrial, etc.) que será representada pela taxonomia; 2) Extração do conjunto de termos ou categorias que representam essa realidade: estabelecer as fontes prioritárias

(fontes pessoais formadas por usuários da web e especialistas no domínio web; fontes documentais, constituídas por documentos representativos dos tipos de conteúdos identificados na fase de planejamento estratégico e taxonomias ou instrumentos de representação do conhecimento existentes. 3) Controle dos termos ou categorias, que compreende a identificação dos diferentes sinônimos e, se for necessário, determinar quais termos são considerados preferidos e quais são considerados não preferidos. Em seguida dar uma forma correta e consistente a todos os termos da taxonomia. O resultado deste processo é o estabelecimento da relação de equivalência entre os termos da taxonomia. 4) Estabelecimento do esquema e estrutura dos termos ou categoria, a partir dos critérios utilizados para dividir e agrupar as categorias e definir o tipo de relação estabelecida entre os agrupamentos de categorias derivadas do esquema de organização. Centelles (2005) considera que a tendência geral tem sido a aplicação de modelos hierárquicos baseados na relação "tipo de" e em árvore com base na relação "parte de". Um terceiro modelo, o facetado, constitui uma alternativa para o ambiente hipertextual, no qual é fundamental a decomposição das diferentes perspectivas a partir das quais um mesmo conceito ou item pode ser observado. Se aprecia também como técnicas tradicionais para desenvolver a estrutura da taxonomia as de *top-down* (descendente) e *bottom-up* (ascendente). O artigo finaliza discutindo a aplicação da taxonomia no desenvolvimento de sistemas de busca de informação, para o qual há diferentes opções: apresentar a estrutura com todas as categorias e relações que as interligam; apresentá-la parcialmente para destacar o conteúdo a partir de critérios temporais ou de uso; apresentar apenas as relações de equivalência, de modo que a taxonomia assume a forma de um anel de sinônimos; e apresentar somente as relações hierárquicas que permite explorar desde as categorias.

O artigo de Hurtado e Losada (2011) procura atingir o objetivo de definir uma estrutura de classificação de riscos, relacionada ao processo de terceirização do desenvolvimento de software. Se situa no gerenciamento de projetos de terceirização de software nos que se destacam suas altas taxas de falhas. Nesse sentido "Uma taxonomia de riscos fornece uma estrutura de classificação importante para o processo de gerenciamento de risco e um ponto de partida para realizar, fácil e eficientemente, o processo de gerenciamento de riscos (Hurtado; Losada, 2011, p. 94). Os autores afirmam dois motivos para a escolha do método. Em primeiro lugar que o desenvolvimento de uma taxonomia de riscos demanda uma grande análise conceitual que os algoritmos não fazem. E, em segundo lugar que o método MECT (*Method for the Construction of a Taxonomy*) tem sido amplamente utilizado para desenvolver a estrutura da taxonomia de riscos para o processo de terceirização do desenvolvimento de software. O MECT se descompõe nas fases de: 1) Planejamento do contexto, do público e do conteúdo: permitirá determinar as características da área do conhecimento que será abordada e para a qual será desenvolvida a taxonomia. Portanto, deve haver um plano explícito para o contexto em que a taxonomia será enquadrada, o público ao qual será dirigida e o conteúdo que compreenderá; 2) Delimitação da área de conhecimento: a preocupação é garantir que a estrutura da taxonomia atenda aos objetivos para que é desenvolvida; 3) Definição de categorias que representam a área de conhecimento: uma área de conhecimento geralmente é representada por diferentes aspectos que satisfazem seus objetivos gerais. Para isso, o método MECT se propõe a analisar informações específicas seguindo as etapas de: a) Definição de fontes e b) Mecanismos de extração das categorias; 4) Estabelecimento das relações entre as categorias: definidas de acordo com a terminologia da temática, considerando os seguintes aspectos: a) identificar os termos representativos da área de estudo e b) todos os termos da taxonomia

podem ser atribuídos de forma correta e consistente, o que dará origem às categorias que irão estruturar a taxonomia.5) Estabelecimento do esquema e estrutura das categorias realizada da seguinte forma: a) definir os critérios usados para dividir e agrupar categorias e b) estabelecer uma técnica que permita a representação gráfica dos componentes da taxonomia. O MECT propõe o uso das técnicas de processamento *Top-Down* e *Bottom-up*. Se considera que a equipe envolvida contará com representantes do campo de conhecimento, sobretudo na etapa de definição de categorias: especialistas da equipe de trabalho, redes e grupos de investigação envolvidos que contribuirão com conhecimento e com experiência profissional e acadêmica para fixar a situação contextual. Como resultado mais destacável se obtém a estrutura base da Taxonomia de Risco de Terceirização de Software, conforme Fig. 1, desenhada em formato de árvore e representada por quatro categorias a partir da aplicação do método aos riscos em terceirização de software. O que leva a concluir que as principais vantagens que oferece essa taxonomia são organizar informações iniciais e fundamentais para a fase de identificação de riscos, com vistas a apoiar a realização efetiva do processo de gerenciamento de riscos.

A dissertação de mestrado de Aganette (2010) se propôs analisar a literatura publicada sobre definições e métodos de elaboração de taxonomias corporativas. O estudo se fez sobre trinta definições e onze métodos encontrados na literatura para construir taxonomias que foram analisados segundo as seguintes categorias: Tipo: origem, uso e elaboração; Essência: atributo, classificação, estrutura e terminologia; e Usos, aplicações e finalidades: potencialidades e gestão da informação. A pesquisa aponta similaridades entre as etapas de construção marcadas nas diferentes propostas analisadas e destaca alguns problemas relacionados à falta de um modelo padrão e à ausência de explicações e procedimentos para a prática de elaboração. Esta limitação é claramente identificada nas metodologias analisadas (Aganette, 2010): as práticas são apenas citadas, não deixando claro como fazer o procedimento específico de cada etapa da taxonomia.

Como resultado principal é alcançado um modelo de construção composto por onze etapas (Aganete, 2010): 1) Definição do domínio de conhecimento; 2) Análise das informações coletadas; 3) Coleta dos termos; 4) Análise dos termos selecionados; 5) Estabelecimento das categorias gerais da taxonomia; 6) Construção dos relacionamentos semânticos; 7) Validação da taxonomia; 8) Definição da forma de apresentação da taxonomia; 9) Definição da tecnologia de suporte; 10) Publicação da taxonomia; 11) Realização de manutenção na taxonomia. Se descrevem cada uma dessas etapas indicando em que consistem e qual é o seu papel na construção da taxonomia, entretanto a proposta não detalha os procedimentos de execução das etapas. A fundamentação teórica da pesquisa compreende os 11 métodos de construção de taxonomia, encontrados na literatura analisada pela autora.

Por sua vez, o artigo de Viana, Aganette e Maculan (2021) intitulado “Metodologia para construção de taxonomias corporativas facetadas” tem como objetivo “sistematizar os procedimentos para construção de taxonomias corporativas, de modo a ressignificar e caracterizá-las como facetadas.” Se pretende fornecer os elementos e atributos necessários para elaborar, desenvolver e manter taxonomias facetadas com foco nas taxonomias corporativas.

Dois pontos merecem destaque na abordagem da estrutura da taxonomia deste estudo. As taxonomias de classificação facetada permitem que o usuário localize, a princípio, o assunto mais amplo e, na sequência, as informações mais restritas. De modo que oferecem vantagens como possibilitar a melhoria da eficiência da recuperação dos conteúdos organizacionais, excluindo as ambiguidades, com a possibilidade de incluir e alterar conceitos sem necessidade de modificar a estrutura da taxonomia. Essa mudança permite que o usuário perce-

ba com facilidade o campo de aplicação devido ao uso de categorias mais segmentadas e que contribuem para a explicitação das informações. Em segundo lugar as pesquisadoras se têm apoiado na asserção de Campos e Gomes (2008) que reitera a necessidade de estabelecer novas relações semânticas para as taxonomias além das tradicionais hierárquicas, muito restritas em suas possibilidades de exploração. Procedimento que se adota no artigo para responder aos requisitos do contexto de aplicação, bem como para integrar as fontes de informação existentes. Os fundamentos teóricos que subsidiaram a proposta compreendem temáticas da organização do conhecimento como Teoria da classificação, Classificação facetada de Ranganathan, Requisitos de modelos conceituais, princípios de construção de taxonomias, entre outros

O artigo adota a metodologia proposta por Aganette (2010) para construir taxonomias corporativas com a ideia de sistematizar os procedimentos de ressignificar e caracterizar as taxonomias como facetadas. Sem esquecer o aporte teórico de outras publicações sobre o tema. A proposta de metodologia de construção de taxonomias corporativas pode ser sintetizada em 11 etapas a saber: 1) Coleta de termos; 2) Estabelecimento das categorias gerais da taxonomia; 3. Definição do domínio de conhecimento; 4) Análise das informações coletadas; 5) Construção dos relacionamentos semânticos; 6) Análise dos termos selecionados; 7) Definição da tecnologia de suporte; 8) Definição da forma de apresentação da taxonomia; 9) Validação da taxonomia; 10) Realização de manutenção na taxonomia e 11) Publicação da taxonomia. A fundamentação teórica para a proposta metodológica se situa na Ciência da Informação e compreende temáticas como classificação facetada; teoria da classificação; requisitos de modelos conceituais, sistemas hipertextuais e modelo de hipertexto com abordagem semântica.

O capítulo Taxonomia do livro “Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais” (Moreiro, 2011) aborda o uso contemporâneo das taxonomias terminológicas e atende ao seu processo de elaboração. Desatacando que, antes de pensar nas etapas de construção, se requer um planejamento detalhado com vistas a entender a sua dimensão aplicativa. Nesse sentido, reforça, ainda a importância de verificar a existência de taxonomia sobre a mesma temática antes de iniciar o processo de construção, visto que a reutilização de taxonomias existentes ou de outros tipos de vocabulários podem ser úteis e ajuda a reduzir ou evitar o retrabalho.

Acerca da estrutura classificatória da taxonomia, o texto destaca alguns princípios importantes para garantir a sua consistência e alcançar a recuperação da informação desejada. Em especial, chama a atenção para alguns critérios gerais que devem ser considerados na hora de determinar o tamanho da estrutura da taxonomia em relação aos termos empregados. Assim como enfatiza que o estabelecimento dos termos escolhidos para representar as categorias devem considerar os princípios de Comunicabilidade, quando devem representar o conceito reconhecido pelos usuários; Utilidade a partir da granularidade suficiente para representar um conteúdo que se mostre num conjunto de documentos; Motivação que induz o usuário desde o uso de termos a continuar navegando no sistema; e Compatibilidade, o uso em referência às atividades e funções da organização correspondente.

O capítulo analisado examina os três métodos básicos para construir taxonomias de Evert Jagerman (2006): do genérico ao específico (método mais tradicional); do específico ao genérico e de aproximação facetada. As informações sobre cada método são agrupadas em passos com uma indicação breve das tarefas correspondentes.

O método do genérico ao específico consiste em sete passos: 1) levantar e registrar o material; 2) reduzir os sinônimos e as variações ao termos preferente e dispor todos eles em

ordem alfabética; 3) elaborar estrutura sistemática prévia; 4) elaborar a primeira versão do esquema; 5) completar a taxonomia com todos os seus termos; 6) comprovar o funcionamento em experimentos de indexação e recuperação; 7) publicar uma versão da taxonomia.

Enquanto o método do específico para o genérico: 1) formar um grupo de expertos no domínio da taxonomia a ser criada; 2) formar uma coleção representativa de documentos já indexados; 3) criar uma coleção de documentos, se estes não estiverem indexados; 4) criar a primeira versão do esquema taxonómico; 5) contar com expertos no domínio para esclarecer o significado de termos ambíguos; 6) avaliar a taxonomia mediante provas de indexação e recuperação.

Por fim, o método da aproximação facetada: 1) criar uma lista de termos que representem conceitos simples; 2) agrupar os termos em categorias; 3) classificar as facetas em um número restrito de categorias fundamentais; 4) usar notações para fixar criar valor classificatório de cada termo dentro da sequência; 5) fixar o significado, mediante definições dos termos mais abrangentes.

O capítulo faz uma proposta de construção sintética a partir dos métodos acima que, em resumo, compreende quatro etapas (Moreiro, 2011): 1) Mineração de dados: identificação dos pontos de acesso que serão incluídos na estrutura classificatória. 2) Análise dos documentos e informações que serão agregadas à taxonomia: por meio de conhecimento dos especialistas. 3) Elaboração da estrutura classificatória da taxonomia: compreende definição do escopo da taxonomia indicando especialidade, processos ou tarefas. Importante definir a superestrutura representativa, formas gráficas de representação e o software a utilizar. 4) Verificação por especialistas: confirmação do trabalho realizado e transferência do conhecimento.

A proposta de Moreiro Gonzáles (2011), formada por cinco etapas contempla a estrutura de uma taxonomia terminológica e se fundamenta nos princípios classificatórios e subsídios teóricos e epistemológicos da Organização do Conhecimento na Ciência da Informação.

O quadro 3 (Apêndice A) apresenta uma síntese das metodologias para construção de taxonomias que foram analisados nesta pesquisa.

## **6 CONSIDERAÇÕES FINAIS**

Após sistematizar os dados acerca dos procedimentos metodológicos das nove publicações analisadas para responder à pergunta de pesquisa, alguns pontos merecem destaque. Em um cenário marcado pela forte presença das tecnologias de informação e comunicação, como propulsoras da criação de soluções metodológicas, de processos para representar e disponibilizar informações, os vocabulários estruturados como as taxonomias desempenham o papel de ferramentas para a organização de informações das instituições de maneira geral. Nesse sentido, as taxonomias como sistemas de organização do conhecimento são criadas para funcionar como esquemas classificatórios de diferentes objetos de conteúdo, potencializadas pelo uso dos termos presentes em textos associados a diferentes tipos de imagens, sons, fórmulas matemáticas e de inúmeros recursos que agregados, demandam por classificação e são fundamentais para a recuperação de informação. Entretanto, é interessante observar que as propostas metodológicas analisadas revelam um padrão similar de procedimentos para o desenvolvimento da estrutura dos componentes de uma taxonomia. Em relação a esse quesito, observou-se que faltam em algumas propostas um detalhamento maior acerca dos pro-

cedimentos que envolvem a estrutura dos termos, do processo de categorização, do uso da classificação facetada e outros elementos acerca das entidades da taxonomia. Outro aspecto a ser mencionado e que fica claro nas propostas metodológicas é a possibilidade de customização das taxonomias. Nesse sentido, as propostas revelam a flexibilidade para adaptar a taxonomia às necessidades específicas do domínio de conhecimento onde a metodologia foi aplicada, considerando as características particulares de cada organização.

Ainda que cientes da ampliação do uso de outros tipos de relações semânticas, conforme já apontado na literatura da Ciência da Informação, a proposta automatizada de Andriani (2017) pode ser considerada como auxílio para a extração, a verificação e a validação das relações hierárquicas por meio do método baseado em coocorrência. Por outro lado, salienta-se a importância do uso de procedimentos tecnológicos a partir do estabelecimento de procedimentos operacionais previamente estabelecidos para organizar a construção dos componentes da estrutura da taxonomia.

Em uma análise sobre o nível de detalhamento dos procedimentos metodológicos do *corpus* da pesquisa verificou-se a presença de propostas com diferentes níveis de descrição do método proposto. Essa situação pode denotar um risco de compressão dos procedimentos de implementação por parte da equipe e centrar excessivamente em decisões imediatas tomadas no momento da realização da tarefa. De maneira geral pode-se afirmar que a construção de uma taxonomia é um projeto complexo, sobretudo as etapas referentes à identificação, estruturação e normalização dos termos e processo de categorização que demandam por atividades de modelagem conceitual. Outro ponto que pode ser destacado é a necessidade de pensar na junção de esforços de profissionais com conhecimento distintos para a construção dos elementos dos componentes estruturais das taxonomias em um projeto de construção de taxonomias. Pode-se inferir que o conhecimento dos profissionais da Biblioteconomia e Ciência da Informação associados à formação dos profissionais com habilidades de lógica, de linguística de desenvolvimento de soluções tecnológicas e matemáticas, além de outros especialistas com formação diversificada e do usuário pode resultar em resultados satisfatórios do ponto de vista da estrutura das taxonomias.

Uma reflexão que se pode propor é acerca da afirmativa de Arudo; Centelles (2005) que as metodologias para construção dos tesouros podem ser usadas como subsídios relevantes para o processo de construção das taxonomias. Naturalmente a afirmativa dos autores se sustenta, visto que a Ciência da Informação dispõe de uma ampla gama de pesquisas, normas técnicas e manuais de construção de tesouros validados pela comunidade internacional. Mas, registra-se aqui a pergunta: o caminho para padronizar uma recomendação do uso dessas metodologias deveria ser percorrido? Essa resposta merece a realização de outros estudos, bem como envolve discussões entre os pesquisadores da área.

Com os apontamentos feitos na seção anterior, entende-se que a pergunta de pesquisa foi respondida e o objetivo atingido. Entretanto, não se pode deixar de mencionar que os resultados alcançados nesta pesquisa, bem como as considerações realizadas estão limitadas às publicações bibliográficas recuperadas no protocolo de busca, realizado e apresentado na seção metodologia. Por último, pode-se inferir que as metodologias apresentadas com mais riqueza de detalhes parecem oferecer uma maior orientação para a execução das etapas referentes à criação dos elementos da estrutura dos componentes das taxonomias. Entretanto, cumpre registrar a importância de ampliar este estudo com vistas a identificar elementos mais concretos acerca da estrutura classificatória e do uso de taxonomias pelos usuários analisando a taxonomia do ponto de vista do usuário. Ainda, é necessário avançar para um levantamento bibliográfico e documental com vistas a estender a análise a publicações de pesquisadores de



outras nacionalidades que também investigam o tema e especificamente a profissionais cuja atividade profissional compreende o desenvolvimento de taxonomias e que podem contribuir identificando os desafios acerca da construção dos componentes da estrutura da taxonomia compartilhando um olhar da realidade acerca do tema.

## REFERÊNCIAS

- AGANETTE, E. C. **Taxonomias corporativas**: um estudo sobre definições e etapas de construção fundamentado na literatura publicada. 104 f. 2010. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.
- ANDRIANI, M. L. **Um método para a construção de taxonomias utilizando a DBPEDIA**. 149 fls. 2017. Dissertação (Mestrado em Engenharia do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina.
- ARUDO, S.; CENTELLES, M. Metodología para el diseño de taxonomías corporativas. **Investigación bibliotecológica**, v. 19, n. 39, Julio/Diciembre, 2005.
- CAMPOS, M.L.A.; GOMES, H.E. Taxonomia e classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais [...]** Salvador: EDUFBA, 2007.
- CARLAN, E.; MEDEIROS, M.B.B. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **RICI: Revista Ibero-americana de Ciência da Informação**, Brasília, v. 4, n. 2, p. 53-73, ago./dez.2011.
- CENTELLES, M. Taxonomías para la categorización y la organización de la información en sitios web [en línea]. **HIPERTEXT.NET, Revista académica sobre documentación digital y comunicación interactiva**, n. 23, 2005.
- GILCHRIST, A.; KIBBY, P.; Mahon, B. **Taxonomies for business**: access and connectivity in a wired World. London: TFPL, 2000.
- GOMES, Hagar Espanha. Taxonomia e a web, construção e uso. <https://www.conexaorio.com/bit/taxonomianaweb.pdf>. 2014.
- HEDDEN, H. **The Accidental Taxonomist**. 2.ed. Nueva Jersey, Estados Unidos de América: Information Today. 2020.
- HEDDEN, H. **The Accidental Taxonomist**. Nueva Jersey, Estados Unidos de América: Information Today. 2010.
- HURTADO, G. P. G.; LOSADA, B. M. Method to construct a taxonomy: basic structure to risks in software outsourcing. **Revista Facultad de Ingenieria-Universidad de Antioquia**, n. 60, p.93 – 101, Sep. 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964, **Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval**. Geneva: The Organization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964, **Information and documentation – Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies**, 2013.

JAGERMAN, E. J. **Creating, maintaining and applying quality taxonomies**. Zoetermeer: E.J. Jagerman, 2006.

MARTÍNEZ, A.; Ristuccia, C.; Pisarello, R.; et. al. Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios Web argentinos. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 106-111, maio/ago. 2004.

MOREIRO GONZÁLEZ, J. A. Taxonomia. In MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais**. Salvador: Edufba, 2011. Cap.6, p.51 – 63.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SOUSA, H. Navegando pelas eras digitais: dos primórdios da Web 1 à revolução da Web 3. 2023 (online). Disponível em <https://malvamarketingdigital.com/2023/09/26/eras-digitais-web3/>. Acesso em: 17 set. 2024.

SUÁREZ SÁNCHEZ, A.; LÓPEZ HERNÁNDEZ, A. Taxonomías digitales: fundamentos teóricos y metodología de construcción. **e-Ciencias de la Información**, v. 10, n. 1, 2020.

VIANA, J. Q.; AGANETTE, E. C.; MACULAN, B. C. M. dos S.. Metodologia para construção de taxonomias corporativas facetadas. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, n. Especial, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37162>. Acesso em: 17 set. 2024.

VITAL, L.P.; CAFÉ, L. M. A. Práticas de elaboração de taxonomias: análise e recomendações. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais [...]** Salvador: ANCIB/UFBA. 2007.

## AGRADECIMENTOS

Ao professor supervisor do estudo de pós-doutorado, à Universidade Carlos III de Madrid pela acolhida recebeu como professor visitante e ao apoio da CAPES: O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

**APÊNDICE A - QUADRO 3 – SÍNTESE DAS ETAPAS DA METODOLOGIAS PARA A CONSTRUÇÃO DE TAXONOMIAS**

1. Martínez, et. al. (2004)	2. Andriani (2017)	3. Suárez Sánchez; López Hernández, (2020)	4. Arudo; Centelles (2005)	5. Centelles (2005)
<p>1) Análise da taxonomia de dos sites, identificando as classes ou táxons de primeiro e segundo nível, ou seja, as classes mais gerais.; 2) Classificação dessas classes em uma das categorias ou facetas fundamentais utilizadas do CRG;3) Agrupamento das classes nas respectivas categorias fundamentais, de acordo com os diferentes itens. 4) Os seis grupos foram comparados para determinar até que ponto as categorias fundamentais do CRG eram comuns às diferentes áreas.</p>	<p>1) Extração de termos a serem utilizados pela taxonomia; 2) Verificação das relações entre os termos extraídos anteriormente. 3) Geração da árvore taxonômica. O método possui duas variações de algoritmo-base, que tratam da forma de se testar a validade de relações taxonômicas: 1) Método baseado em coocorrências dentro do corpus; e 2) Método baseado em coocorrências dentro do currículo:</p>	<p>1) Determinação da realidade: é a determinação de entidades, áreas de conhecimento, setores industriais, etc. que será representado pela taxonomia (Centelles, 2005). 2) Estabelecimento de termos: envolve obter informações sobre quais podem ser incluídos e, ainda, avaliar para decidir quais devem ser incluídos (Hedden, 2010); 3) Normalização de termos: consiste em duas tarefas: determinar os termos preferidos versus aqueles que não o são e dar uma forma correta e consistente a todos os termos da taxonomia (Centelles, 2005). 4) Criação de categorias: estruturação da taxonomia, considerando relações de equivalência, hierarquia, associação e semântica (Hedden, 2010). 5) Estruturação em um programa: quando a taxonomia se torna muito complexa contém várias centenas de termos, deve-se buscar software de taxonomia (Hedden, 2010).</p>	<p>1) Planejamento estratégico a) Análise de contexto Análise de público; b) Análise de conteúdo; c) Definição de taxonomia: (c1) identificação dos critérios das facetas de conteúdo; (c2) extração lexical; (c3) controle lexical; e (c4) o desenvolvimento da estrutura da taxonomia. 2) Construção da estrutura; 3) Categorização de conteúdo; 4) Candidatura e submissão, ciclo de avaliação e melhoria contínua e,5) Divulgação.</p>	<p>1) Delimitação da realidade. 2) Extração do conjunto de termos ou categorias que representam essa realidade. 3) Controle terminológico de termos ou categorias. 4) Estabelecimento do esquema e estrutura organizacional dos termos ou categorias</p>

6. Hurtado; Losada (2011)	7. Aganette (2010).	8. Viana; Aganette; Maculan(2021)	9. Moreiro (2011)
<p>Método MECT (<i>Method for the Construction of a Taxonomy</i>) com as etapas de:</p> <ol style="list-style-type: none"> <li>1) Planejamento do contexto, do público e do conteúdo.</li> <li>2) Delimitação da área de conhecimento</li> <li>3) Definição de categorias que representam a área do conhecimento.</li> <li>4) Estabelecimento da relação entre as categorias</li> <li>5) Estabelecimento do esquema e estrutura das categorias.</li> </ol>	<ol style="list-style-type: none"> <li>1) Definição do domínio de conhecimento;</li> <li>2) Análise das informações coletadas;</li> <li>3) Coleta dos termos.</li> <li>4) Análise dos termos selecionados;</li> <li>5) Estabelecimento das categorias gerais da taxonomia;</li> <li>6) Construção dos relacionamentos semânticos;</li> <li>7) Validação da taxonomia;</li> <li>8) Definição da forma de apresentação da taxonomia;</li> <li>9) Definição da tecnologia de suporte;</li> <li>10) Publicação da Taxonomia</li> <li>Realização de Manutenção na taxonomia</li> </ol>	<ol style="list-style-type: none"> <li>1) Coleta de termos;</li> <li>2) Estabelecimento das categorias gerais da taxonomia;</li> <li>3) Definição do domínio de conhecimento;</li> <li>4) Análise das informações coletadas;</li> <li>5) Construção dos relacionamentos semânticos;</li> <li>6) Análise dos termos selecionados;</li> <li>7) Definição da tecnologia de suporte;</li> <li>8) Definição da forma de apresentação da taxonomia;</li> <li>9) Validação da taxonomia;</li> <li>10) Realização de manutenção na taxonomia e</li> <li>11) Publicação da taxonomia.</li> </ol>	<ol style="list-style-type: none"> <li>1) Mineração de dados: identificação dos pontos de acesso que serão incluídos na estrutura classificatória;</li> <li>2) Análise dos documentos/informações que serão agregadas à taxonomia: por meio de conhecimentos dos especialistas;</li> <li>3) Elaboração da estrutura classificatória da taxonomia: compreende definição do escopo da taxonomia indicando especialidade, processos ou tarefas. Importante definir a superestrutura representativa, formas gráficas de representação e o software a utilizar;</li> <li>4) Verificação por especialistas: confirmação do trabalho realizado e transferência do conhecimento.</li> </ol>

Fonte: Dados da pesquisa (2024).