

O uso de taxonomias para o enriquecimento semântico de objetos

The use of taxonomy for the semantic enrichment of objects

Júnio Soares Dias (1), Claudiane Emanuele Nazário (2), Célia da Consolação Dias (3)

Universidade Federal de Minas Gerais (UFMG), Av. Pres. Antônio Carlos, 6627 – Pampulha-MG, Brasil, (1) juniodias@ufmg.br. (2) cen2001@ufmg.br (3) celiadias@eci.ufmg.br.

Resumo

A literatura da Ciência da Informação descreve a necessidade de estudos sobre o enriquecimento semântico de Sistemas de Organização do Conhecimento. Esta investigação apresenta uma análise primária de duas pesquisas em andamento sobre temáticas pertencentes à Organização do Conhecimento. O objetivo deste estudo foi iniciar a análise do estado da arte sobre a taxonomia, o enriquecimento semântico de objetos e o uso da taxonomia destinada ao enriquecimento semântico. Para isto, optamos por utilizar o Mapeamento Sistemático como metodologia de pesquisa nos últimos cinco anos de publicações em quatro bases de dados de pesquisas científicas. Os resultados obtidos foram uma coleta de dados, uma análise quantitativa e qualitativa. A análise quantitativa obtida representa as tendências da temática de pesquisa para o período investigado. Já a análise qualitativa representa o levantamento inicial do estado da arte para o estudo em questão. Concluiu-se que esses estudos obtidos sobre os Sistemas de Organização do Conhecimento e, principalmente, a taxonomia destinada ao enriquecimento semântico são incipientes e o seu desenvolvimento pode contribuir para a Ciência da Informação e outras áreas do conhecimento.

Palavras-chave: Taxonomia. Enriquecimento Semântico. Mapeamento Sistemático.

Abstract

The Information Science literature describes the need for studies about the semantic enrichment of Knowledge Organisation Systems. This research presents a primary analysis of two ongoing investigations on a theme belonging to the Knowledge Organisation. Our goal was to begin the analysis of the state of the art on taxonomy, the semantic enrichment of objects and the use of taxonomy intended for semantic enrichment. For this, we are close to use Systematic Mapping as a research methodology in the last five years of publications in four scientific research databases. The results obtained were a data collection, a quantitative and qualitative analysis. The quantitative analysis obtained represents the trends of the research theme for the investigated period. The qualitative analysis represents the initial survey of the state of the art for the study in question. We conclude that these studies on Knowledge Organisation Systems and especially the taxonomy intended for semantic enrichment are incipient and their development can contribute to Information Science and other knowledge areas.

Keywords: Taxonomy. Semantic Enrichment. Systematic Mapping.

Resumen

La literatura de Ciencias de la Información describe la necesidad de estudios sobre el enriquecimiento semántico de los Sistemas de Organización del Conocimiento. Esta investigación presenta un análisis primario de dos investigaciones en curso sobre temas pertenecientes a la Organización del Conocimiento. El objetivo de este estudio fue iniciar el análisis del estado del arte sobre la taxonomía, el enriquecimiento semántico de los objetos y el uso de la taxonomía destinada al enriquecimiento semántico. Para ello, hemos optado por utilizar el Mapeo Sistemático como metodología de investigación en los últimos cinco años de publicaciones en cuatro bases de datos de investigación científica. Los resultados obtenidos fueron una recopilación de datos, un análisis cuantitativo y cualitativo. El análisis cuantitativo obtenido representa las tendencias de la temática de investigación para el periodo investigado. Ya el análisis cualitativo representa la encuesta inicial del estado del arte para el estudio en cuestión. Concluimos que estos estudios sobre los Sistemas de Organización del Conocimiento y, principalmente, la taxonomía destinada al enriquecimiento semántico son incipientes y su desarrollo puede contribuir a la Ciencia de la Información y otras áreas del conocimiento.

Palabras clave: Taxonomía. Enriquecimiento Semántico. Mapeo Sistemático.

1 Introdução

O crescimento desordenado de dados e informações na *web* tem se tornado um tema recorrente na literatura da área de Biblioteconomia e Ciência da Informação (BCI). Existem repositórios de diferentes domínios do

conhecimento disseminados diariamente sem adotar qualquer padrão ou controle terminológico, o que dificulta sua interpretação por sistemas informatizados.

O enriquecimento semântico é um processo utilizado para atribuir maior significado aos dados e metadados,

através do uso da semântica atribuída por vocabulários pré-existent, sinônimos e informações de proveniência, visando facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas (LIRA, 2014). Conforme apontado na pesquisa do referencial teórico, seção 2, há poucas discussões na BCI referentes ao processo de enriquecimento semântico, sobretudo relacionados à utilização de Sistemas de Organização do Conhecimento (SOC) tais como: taxonomias e ontologias para realizar o enriquecimento semântico e facilitar o processamento dos dados por homens e máquinas.

Na BCI, existem diversos esquemas de representação do conhecimento próprios para a representação da informação e que cumprem a função de comunicação entre os sistemas de informação. Nesta categoria de esquemas, podem ser citados as ontologias, os mapas conceituais e as taxonomias. Todos esses conceitos têm sua origem em outras áreas de conhecimento. As taxonomias têm sua origem na Biologia, e em relação à etimologia, o termo tem sua origem no grego *Táxis* que significa ordem e *Nomos* é igual à lei, norma ou regra. O conceito de taxonomia está relacionado com classificação e com sistemas de classificação. Para a compreensão do significado do termo resgata-se a sua definição presente na Norma ANSI/NISO Z39.19-2005 [1]. De acordo com a norma, a taxonomia é um conjunto de termos de um vocabulário controlado, organizados numa estrutura hierárquica, onde cada termo de uma taxonomia se encontra em mais do que uma relação com outros termos de outra taxonomia.

Esse artigo, cujo objetivo principal é avaliar as contribuições das taxonomias para o enriquecimento semântico de dados, é parte de duas pesquisas de doutorado mais abrangentes, ainda em desenvolvimento. Sendo assim, em busca do estado da arte sobre a taxonomia e o enriquecimento semântico, a questão de pesquisa para atingir esse objetivo estabelecido é: Em que medida o uso da taxonomia pode favorecer o enriquecimento semântico de objetos?

O presente artigo está dividido em cinco seções, a saber: A primeira seção contém os aspectos introdutórios e objetivos da pesquisa. A segunda seção é composta da exploração preliminar da literatura, contendo o referencial teórico, em suporte ao estudo apresentado. A seção três apresenta a metodologia utilizada para elaboração deste estudo. Já a seção quatro apresenta os resultados encontrados. Na última seção, são descritas as considerações finais e na sequência estão as referências bibliográficas usadas para a construção deste artigo.

2 Referencial teórico

Ao realizar uma revisão da literatura sobre o uso da taxonomia para o enriquecimento semântico, foi

permitido contemplar os principais conceitos que circundam o objeto deste projeto de investigação. Hjørland (2009) esclarece que os SOC, contextualizados pela própria área da BCI, são sistemas que podem organizar os conceitos pertencentes a um determinado domínio e as suas relações semânticas. Isto é, os SOC visam à organização da produção intelectual humana e são utilizados para representar e recuperar a informação (BRASCHER; CARLAN, 2010; DIAS, C. C. *et al.*, 2022).

Os SOC podem ser utilizados em diversas áreas do conhecimento, em inúmeras aplicações. Entre esses SOC e áreas do saber, podem-se citar como exemplos a taxonomia para a Ciência da Computação, nos processos da engenharia de requisitos (FAGUNDES *et al.*, 2020), para o Direito, nos argumentos jurídicos (LIMA; GOMES NETO, 2018), para a teoria de linguagem (AUSTIN, 2021) e para pesquisas científicas, nas temáticas da taxonomia corporativa e facetada (LOPES; AGANETTE; MACULAN, 2020), e na interoperabilidade (DIAS, C. C. *et al.*, 2022).

Ao considerar as pesquisas da Ciência da Informação, com foco na taxonomia e no enriquecimento semântico, podem-se citar as pesquisas de Maculan e Lima (2011), Aganette (2010), Aquino, Carlan e Brascher (2009), Barbosa (1972) e Ranganathan (1960) especificamente para a taxonomia. Além dos estudos de Lóscio, Burle e Calegari (2017), Lira (2014) e Sorrentino *et al.* (2013), que abordam o enriquecimento semântico.

Para a taxonomia como referencial teórico, a pesquisa de Barbosa (1972, p. 73) apresenta a necessidade de se “[...] considerar um documento, não apenas por seu aspecto físico mas também por seu conteúdo [...]”. Conforme apresentado pela autora, o aspecto do conteúdo levou à teoria de múltiplos aspectos ou facetas desenvolvida por Ranganathan (1960). Esse estudo aborda a taxonomia, que é constituída de um modelo de representação do conhecimento, discutido em Hjørland (1998).

A taxonomia constitui um modelo de representação do conhecimento, tomada de empréstimo pela Ciência da Informação, com origem na Biologia, como abordado no estudo conceitual de Aquino, Carlan e Brascher (2009). Para esses autores, o termo taxonomia

[...] derivou-se de um dos ramos da Biologia que trata da classificação lógica e científica dos seres vivos, fruto do trabalho do médico e botânico sueco Carolus Linnaeus, com sua *Biologia Sistemática* (AQUINO; CARLAN; BRASCHER, 2009, p. 204).

A associação de ideias de uma navegação facetada facilita a recuperação de informações, denominada de paradigma de navegação facetada (MACULAN; LIMA, 2011). As autoras afirmam que taxonomias proporcionam uma representação do relacionamento semântico dos objetos envolvidos na classificação.

A pesquisa de Aquino, Carlan e Brascher (2009) sobre taxonomia descreve que a informação será acessível quando estiver organizada de maneira a facilitar a busca desejada. A literatura apresenta diversas aplicações para a taxonomia. Uma destas aplicações para taxonomias é a navegacional, que destina-se a buscar a organização da informação conforme o seu público-alvo, isto é, uma forma flexível de utilização.

A proposta apresentada por Maia, Lima e Maculan (2017) descreve uma taxonomia dos tipos de relações semânticas para a organização e para a representação do conhecimento mediante uma revisão da literatura. Nesta investigação, as autoras descrevem os tipos de estruturas taxonômicas, como a corporativa e a facetada.

Para Lóscio, Burle e Calegari (2017), o enriquecimento semântico de dados é um conjunto de processos que podem ser usados para aprimorar, refinar ou melhorar os dados brutos ou previamente processados, contribuindo para tornar esses dados um ativo valioso para quase todos os negócios ou empresas. Uma proposta para o desenvolvimento de sistemas que apoia o processamento de dados brutos para a *web* é a taxonomia navegacional.

Portanto, conforme descrito por Aganette (2010), é necessário investigar as estruturas de taxonomias, tais como as taxonomias navegacionais. E por meio desses estudos identificar meios de utilizar esses instrumentos no processo de enriquecimento semântico. Segundo Sorrentino *et al.* (2013), a associação explícita entre um ou mais significados é uma ferramenta chave para o enriquecimento semântico. Utilizar o relacionamento estabelecido na taxonomia permitiria não só aumentar as possibilidades de pesquisa, mas facilitar a recuperação desses dados por sistemas automatizados.

O enriquecimento semântico destina-se a inserir conceitos a partes de um texto, de modo a criar uma estrutura para busca automática de seu significado (THAKKER *et al.*, 2012). Adotar taxonomias para ampliar o significado desses conceitos permitiria aos usuários interligar esses recursos por meio das relações existentes entre os mesmos e encontrar novos recursos. Contudo, as possibilidades de uso das taxonomias no processo de enriquecimento semântico não estão devidamente explicitadas na literatura da BCI.

Conclui-se, então, que, conforme descrito por Lopes, Aganette e Maculan (2020), “[...] a produção dos pesquisadores da área mostra-se muito dispersa e limitada a poucos trabalhos, [...]”. Assim, conforme as autoras, existe uma lacuna de pesquisas para uma área que busca consolidar-se como campo de pesquisa de instrumentos de representação da informação e do conhecimento.

3 Metodologia

A metodologia desenvolvida para a realização desta pesquisa priorizou o modo qualitativo para estabelecer o estado da arte da área da Ciência da Informação sobre o uso da taxonomia para o enriquecimento semântico de objetos. Para atingir este objetivo, optou-se pelo uso do Mapeamento Sistemático (PETERSEN *et al.*, 2007), ou seja, um levantamento de teorias, conceitos, modelos e métodos associado ao tema da pesquisa, conforme publicado na literatura.

Esse mapeamento exploratório possibilitou o aprofundamento em uma área do conhecimento e na temática de pesquisa, por meio de um processo constituído de cinco etapas, descritas a seguir.

3.1 Mapeamento Sistemático

O Mapeamento Sistemático (MS) foi realizado por meio de buscas a artigos científicos, livros, teses e dissertações que versam sobre o tema desta investigação, durante o período de janeiro a março de 2023. As etapas para o MS, conforme descrito por Petersen *et al.* (2007), são: 1 - estabelecer as questões de pesquisa; 2 - realizar as buscas em bases científicas; 3 - realizar a seleção do material (estudos) de interesse; 4 - efetuar o refinamento pelas palavras-chave e os resumos e 5 - extrair os dados e análises.

Na primeira etapa, são estabelecidas as questões de pesquisa, conforme Petersen *et al.* (2007), regularmente, deseja-se mapear as frequências de publicação ao longo do tempo para ver as tendências de uma investigação. Esses objetivos são refletidos nas questões de pesquisa, descritas na introdução, que foram estabelecidas com base no objetivo e possibilita delimitar o escopo do MS.

Ao realizar a segunda etapa, buscas em bases científicas, foram conduzidos estudos, que são pesquisas por fontes de dados científicos. Baseado na questão de pesquisa, é definido e executado um método para a consulta e recuperação destas fontes de informações.

As consultas foram realizadas em bases nacionais, internacionais e interdisciplinares nos últimos cinco anos. As bases estabelecidas para atingirem o objetivo estabelecido são: Repositório Institucional da Universidade Federal de Minas Gerais (RI/UFMG) [2], *Springer Link* [3], *SciELO* [4] e *Wiley Online Library* [5]. Esta consulta foi baseada em termos de busca. Uma boa forma de criar termos de busca (*strings*) é com o uso da *population, intervention, comparison, outcome and context* (PICOC) (KITCHENHAM; CHARTERS, 2007; PETTICREW; ROBERTS, 2005).

Para a seleção do material de interesse, terceira etapa, foram estabelecidos os critérios de inclusão e exclusão (PETERSEN *et al.*, 2007). Estes são usados para excluir estudos que não são relevantes para responder às questões de pesquisa. Na quarta etapa, uma análise por palavras-chave foi estabelecida pela aplicação das

strings e pelas leituras dos resumos, introduções, resultados e conclusões para a classificação dos materiais.

Para finalizar, a quinta etapa se destina a realizar a análise dos estudos identificados como fonte de dados para a investigação. Este levantamento possibilitou uma classificação do material obtido pelas etapas anteriores e é um método sistemático para selecionar apenas os artigos relevantes para o estudo em questão.

Após estabelecimento da questão de pesquisa, descrita como objetivo na seção 1, foram realizadas as buscas conforme os critérios descritos em 3.1.1 e os critérios de inclusão e exclusão em 3.1.2.

3.1.1 Critérios de busca

A estratégia de busca de pesquisas relacionadas iniciou com o estabelecimento das palavras-chave. Estas foram utilizadas de forma combinada ao utilizar operadores lógicos para uma pesquisa experimental com base no objetivo traçado. Os operadores utilizados foram o “AND” para o agrupamento dos conceitos e o “OR” para pesquisas combinadas com o mesmo conceito. O primeiro passo realizado foi estabelecer as palavras-chave para a língua inglesa (*Taxonomy*, *Semantic enrichment*) e portuguesa (Taxonomia, Enriquecimento semântico).

No segundo passo, para cada palavra-chave foi formado um conceito principal de busca para os artigos científicos e outros estudos relacionados. Os conceitos estabelecidos para a língua inglesa são: *Taxonomy*: *taxonomy*; *Semantic enrichment*: *semantic*, *semantic enrichment*. Já para a língua portuguesa, temos os seguintes conceitos: Taxonomia: *taxonomia*; Enriquecimento semântico: *semântico*, *semântica*, *enriquecimento semântico*.

Além disso, nos casos solicitados foi estabelecida a área de conhecimento das buscas como "Ciências Sociais Aplicadas" para a língua portuguesa e "*Applied Social Sciences*" para a língua inglesa. Na base de dados do RI/UFMG, foi estabelecida, também, a busca para as investigações desenvolvidas para programas de pós-graduação, como, por exemplo, o Programa de Pós-Graduação em Gestão e Organização do Conhecimento, isto devido à ausência da opção de seleção da área do conhecimento.

Como exemplo deste critério de busca é descrita, a seguir, a combinação dos conceitos das palavras-chave em uma *string* de busca para as bases de pesquisa *Springer Link* e a *SciELO*: '(*taxonomy AND (semantic enrichment OR semantic)*) OR (*taxonomia AND (enriquecimento semântico OR semântico OR semântica)*)'. Outro exemplo, para a base de pesquisa *Wiley Online Library*, é a seguinte *string*: '*for "taxonomy" in Abstract and "semantic" in Abstract and "taxonomy" in Keywords and "semantic" in Keywords*'.

3.1.2 Critérios de inclusão e exclusão

Os resultados obtidos pela busca foram filtrados na etapa três do MS. Nesta etapa, seleção do material (estudos) de interesse, foi realizada a seleção das pesquisas de interesse, atividade orientada por uma Doutora em Ciência da Informação e executada por dois alunos de doutorado. Esta seleção seguiu os critérios de inclusão e exclusão para o aceite dos estudos.

Foi considerado como critério de inclusão para as pesquisas de interesse os estudos estabelecidos com o objetivo de apresentar pesquisas sobre a taxonomia e/ou a semântica ou o enriquecimento semântico de objetos. Já os critérios de exclusão são: pesquisas que não possuem como assunto a taxonomia e/ou a semântica ou o enriquecimento semântico; pesquisas incompletas apresentadas como trabalhos com menos de quatro páginas; estudos que não estejam em português ou inglês e pesquisa já publicada e contabilizada em outro periódico ou conferência.

4 Resultados

As pesquisas selecionadas foram filtradas pelas palavras-chave, *string* de busca nas bases de dados, e, posteriormente, refinadas pelo conteúdo contido no estudo, etapa quatro do MS. Uma descrição quantitativa dos dados coletados, refinados e analisados nesta etapa, encontra-se na subseção 4.1. A subseção 4.2, revisão da literatura, apresenta a extração e a análise qualitativa dos dados, etapa cinco do MS.

4.1 Dados coletados

Após a coleta de estudos, etapa 3 do MS, que estavam alinhados com o objetivo desta pesquisa, realizou-se o refinamento pelas palavras-chave, resumos e leitura das introduções, resultados e conclusões (pela aplicação de três filtros). Os estudos foram quantificados e os valores obtidos encontram-se descritos na Tabela 1. Esta tabela apresenta nas linhas os nomes das bases de pesquisas utilizadas e nas colunas o identificador dos devidos "Filtros 1, 2 e 3". Sendo que, cada filtro apresenta a quantidade de estudos obtidos por base.

A coluna “Filtro 1” apresenta a quantidade de estudos obtidos pelas *strings* de busca nas bases de dados. Para cada uma dessas fontes de informações científicas foi realizada uma pesquisa com duas *strings* de buscas, conforme especificação das palavras-chave, conceitos de busca, e respeitando a forma de manipulação das quatro bases escolhidas para este estudo.

O quantitativo de estudos para cada uma das bases de dados, após a aplicação do filtro de leitura dos títulos e resumos, ao utilizar do critério de inclusão, encontra-se na coluna “Filtro 2” dessa Tabela 1. Por fim, a coluna “Filtro 3” apresenta o quantitativo final que foi obtido de cada base após a leitura da introdução, resultados e/ou conclusão de cada estudo.

Este último filtro considera o critério de exclusão estabelecido, descrevendo o quantitativo de estudos analisados na quinta etapa do MS. O baixo número de resultados de pesquisas, 67 estudos, já era esperado, isto devido ao período de pandemia da COVID-19 (hipótese identificada na análise dos dados a seguir) e, principalmente, a especificidade do objeto deste nosso estudo.

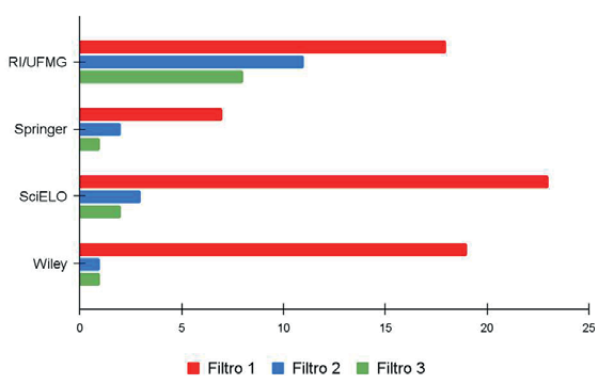
Tabela 1. *Quantitativos de estudos obtidos*

Bases	Filtro 1	Filtro 2	Filtro 3
RI/UFMG [2]	18	11	8
Springer [3]	7	2	1
SciELO [4]	23	3	2
Wiley [5]	19	1	1
Total	67	17	12

Fonte: Elaborada pelos autores (2023).

Pode-se verificar na Tabela 1 que dos 67 estudos, apenas 12 foram considerados alinhados à temática. Esse total de estudos selecionados no "Filtro 3" representa 17,9% das pesquisas obtidas no primeiro filtro. Ao analisar estes estudos selecionados por ano, após a aplicação dos três filtros, obteve-se os Gráficos 1 e 2. O Gráfico 1 apresenta a quantidade de estudos encontrados por base de dados em cada momento de aplicação desses três filtros. Pode-se constatar que, apesar da busca inicial com *strings* ter obtido um número de resultados com média simples de 16,8 estudos para as quatro bases, o aproveitamento das pesquisas ao considerar a temática foi pequeno, uma média simples de 3 estudos.

Gráfico 1. *Quantidades de estudos por filtro*



Fonte: Elaborado pelos autores (2023).

Ao analisar o Gráfico 1, conclui-se que, para a busca estabelecida, as bases de dados de pesquisas científicas com melhor aproveitamento percentual foram as do RI/UFMG e *Springer*. Na base do RI/UFMG, o aproveitamento percentual foi de 44,4% e para a *Springer*, ocorreu um aproveitamento de 14,3%. Para a

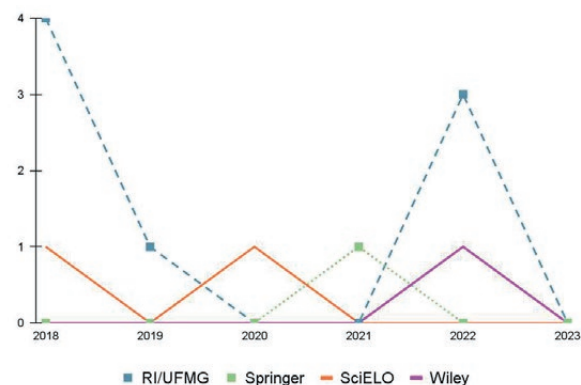
SciELO, ocorreu o aproveitamento percentual de 8,7% e na base *Wiley*, foi obtido o aproveitamento de 5,3%.

Já o Gráfico 2 apresenta a quantidade de publicações de estudos obtidas nas quatro bases de dados durante o período de 2018 a fevereiro de 2023. Percebeu-se o aproveitamento de mais estudos nas fontes da RI/UFMG, SciELO e *Wiley* em 2018 e 2022. Sendo que, como destaque, no ano de 2018, o processo de busca e a aplicação dos filtros resultou no total de 5 estudos e em 2022, este processo obteve 4 estudos.

Deve-se considerar, conforme Tabela 1, que em números absolutos as bases do RI/UFMG e SciELO representam o melhor retorno, com respectivamente 8 e 2 estudos, após a aplicação do "Filtro 3". Atribuímos este resultado expressivo no aproveitamento de estudos na base da RI/UFMG a elementos isolados da temática de pesquisa, como, por exemplo, trabalhos ligados somente ao SOC, pertencente ao objetivo de nosso estudo, a taxonomia.

Percebe-se também na análise do Gráfico 2 que a produção científica para a temática diminuiu para o período de 2019 a 2021, uma possível explicação seria a pandemia de COVID-19, pois esta ocorreu no mesmo período.

Gráfico 2. *Quantidades de estudos das bases de dados*



Fonte: Elaborado pelos autores (2023).

Além dos elementos quantitativos apresentados nesta subseção, a análise dos dados coletados proporcionou o Quadro 1 apresentado no Apêndice A. Esse quadro descreve os estudos obtidos pelo "Filtro 3" para suas respectivas temáticas de pesquisa e as interpretações sobre os objetivos traçados pelos autores. Deve-se observar que entre os resultados verificou-se que há artigos e trabalhos de pesquisas mais complexos e extensos, como dissertações e teses.

Estes estudos abordam de forma isolada a taxonomia, a representação semântica e o enriquecimento semântico, em suas devidas particularidades. Porém, nenhum dos estudos encontrados apresentou uma pesquisa em profundidade sobre a taxonomia e sua relação com o enriquecimento semântico de objetos. Estes fatos

quantitativos, descritos anteriormente, e a análise qualitativa, do Quadro 1, comprovam a necessidade de ampliar os estudos nessa temática de pesquisa estabelecida. Uma análise qualitativa inicial desse Quadro 1 encontra-se descrita na subseção 4.2 - Estado da arte.

4.2 Estado da arte

A aplicação do MS resultou em publicações cujas análises apontaram para o estado da arte na literatura para os últimos cinco anos, em quatro bases de dados de pesquisas científicas. Isto permitiu estabelecer um retrato da temática estabelecida e contemplar os principais conceitos que circundam o objeto deste projeto de investigação. Assim, os estudos contemporâneos descritos, a seguir, contextualizam uma análise primária para o objetivo estabelecido, o uso da taxonomia para o enriquecimento semântico de objetos.

A análise do estado da arte foi baseada nas pesquisas obtidas, que estão resumidas no Quadro 1, entre essas, como exemplos, descrevemos os estudos destinados aos SOC (LOPES; AGANETTE; MACULAN, 2020), a processos e à representação semântica (LISKA, 2018; OLIVEIRA, 2018). Além disto, como destaque, apresentamos os estudos da taxonomia (ANDERSON *et al.*, 2022; SANTOS, 2018; MENDES, 2022; VIANA, 2022) e do enriquecimento semântico (NAZÁRIO e DIAS, 2018).

Uma pesquisa que abordou a taxonomia foi a de Mendes (2022). Neste estudo, a taxonomia é apresentada como um

[...] processo de classificação e agrupamento categorizado, que define-se como organização, ou seja, partindo de um assunto cria-se classes gerais que se dividem em classes e subclasses, hierarquicamente, formando uma lista de assunto estruturada” (MENDES, 2022, p. 44).

Neste contexto, temos então a taxonomia corporativa e a facetada. Lopes, Aganette e Maculan (2020) descrevem que a taxonomia corporativa pode existir em distintos ambientes na organização, pois ela

[...] estrutura e organiza o conhecimento registrado em poder das organizações, organiza produtos e serviços, e se constitui como um mecanismo de apoio à gestão. Ela é construída de modo a atender a um público particular, servindo de interface sistema-usuário por meio da navegação (LOPES; AGANETTE; MACULAN, 2020, p. 164).

Ainda segundo Lopes, Aganette e Maculan (2020), ressalta-se a teoria da classificação facetada como aporte teórico para a construção de taxonomias:

A taxonomia facetada emprega os princípios da Teoria da Classificação Facetada, desenvolvida por Ranganathan (1967), que foi criada com o objetivo de auxiliar na organização de materiais bibliográficos (LOPES; AGANETTE; MACULAN, 2020, p. 164).

As autoras apontam também as vantagens do uso de classes e subclasses de assunto para a recuperação da informação:

Na construção de instrumentos de representação da informação e do conhecimento, a teoria fornece cinco categorias mais abstratas de mapeamento do domínio representado, formando-se múltiplas hierarquias de classes (facetadas) e subclasses (subfacetadas) que exprimem as diferentes dimensões desse dado domínio. Esse modo de classificação permite que o usuário encontre primeiro o assunto mais amplo, e, através dele, chegue à informação mais restrita. Essa navegação torna mais acessível e prática a recuperação da informação por parte do usuário (LOPES; AGANETTE; MACULAN, 2020, p. 164-165).

Duas pesquisas que também tratam com mais profundidade, de forma contemporânea, a taxonomia corporativa e a taxonomia facetada são as de Mendes (2022) e a de Viana (2022). A pesquisa de Mendes (2022) investiga a taxonomia corporativa e facetada. Este estudo concluiu que para iniciar uma pesquisa nesta temática é necessário verificar o foco nas taxonomias corporativas, como vocabulário controlado, para a representação da informação, quando se pesquisa um ambiente organizacional. Esta última autora também conclui que é necessário

[...] utilizar outro insumo tecnológico para o desenvolvimento e até mesmo para o aprimoramento do uso da taxonomia corporativa facetada, como, por exemplo, algum sistema informacional [...] (MENDES, 2022, p. 78).

A investigação de Viana (2022) propõe a taxonomia facetada em ambientes corporativos, considerando-a como instrumento na gestão e na organização da informação. Além disso, Viana (2022) discute os benefícios do uso da taxonomia e da classificação, possibilitando confirmar os resultados dos estudos de Aganette (2010) ao evidenciar que as taxonomias têm representado soluções para encontrar informações em diversos ambientes.

As pesquisas de Viana (2022) e Mendes (2022) indicam que a Teoria da Classificação Facetada, conforme o contexto de necessidades, é uma solução para a melhoria dos sistemas de classificação corporativos, considerando que existem várias possibilidades de classificar assuntos de documento. A citada pesquisa conclui também que, apesar da escassez de estudos, a literatura da Ciência da Informação indica a necessidade de utilizar os princípios classificatórios explícitos para o desenvolvimento de taxonomias.

Já sobre a *web* semântica e a interoperabilidade semântica, duas pesquisas contemporâneas com profundidade que foram encontradas são as de Firmino (2019) e Emygdio (2021). Contudo, apesar da importância das pesquisas para os SOC e do relacionamento com a temática estabelecida, estes estudos não abordam diretamente o enriquecimento semântico.

O enriquecimento semântico foi discutido na pesquisa de Nazário (2018) e em Nazário e Dias (2018), que apresentou uma proposta metodológica para avaliar o enriquecimento semântico de objetos publicados na *web* através do *Linked Data* [6]. Neste estudo, foi desenvolvida uma Matriz de Técnicas e recursos para o enriquecimento semântico de objetos (Matriz Treso), que foi utilizada na avaliação dos modelos de dados *Europeana Data Model* (EDM) [7] e *Bibliographic Framework Initiative* (BIBFRAME) [8]. A pesquisa também apontou a necessidade de aplicação da matriz a outros modelos de dados, bem como das possibilidades de uso da metodologia proposta para criar novos modelos de dados enriquecidos semanticamente.

Já o estudo de Maia (2018) apresentou um modelo de extração de relações semânticas para a representação do conhecimento de documentos acadêmicos no contexto do idioma português brasileiro, implementado em um sistema *web* chamado Semantizar, que realizou o enriquecimento semântico da amostra pesquisada.

Importante destacar que apesar da relevância do enriquecimento semântico para a recuperação e organização do conhecimento no contexto da BCI, foram poucas as pesquisas obtidas para o período de 2018 a fevereiro de 2023 nas bases selecionadas que abordam esse enriquecimento semântico. Os filtros de busca recuperaram, também, artigos sobre relacionamento semântico, semântica e taxonomia. Portanto, esta pesquisa é um estudo em forma de primeira análise sobre os dados coletados, e não foi possível contextualizar de forma completa, nem tão pouco descrever o estado da arte desejado, ao considerar o universo pesquisado em seu contexto.

5 Considerações finais

Este artigo apresenta a análise inicial dos dados que estão sendo coletados para duas pesquisas, que proporcionarão trabalhos com mais profundidades nas temáticas estabelecidas. Podemos concluir que, conforme dados quantitativos e afirmado por Lopes, Aganette e Maculan (2020), existe uma carência de pesquisas sobre os SOC e essa temática, fato comprovado pelo número de pesquisas encontradas.

Ressalta-se que tanto a temática do enriquecimento semântico quanto a de taxonomias e suas possibilidades de utilização para favorecer o processo de enriquecimento semântico demandam mais estudos, tendo em vista que na amostra analisada não foi encontrado nenhum resultado que abordasse esses temas de modo conjunto. Isto é, nenhum dos estudos encontrados apresentou uma pesquisa em profundidade sobre a taxonomia e sua relação com o enriquecimento semântico de objetos. A importância do tema pressupõe a necessidade de seu aprofundamento, sobretudo pelo fato cada vez mais constante da disseminação de conteúdos e informações sem qualquer controle na *web*.

Assim, estudos futuros precisam ser desenvolvidos com o objetivo de identificar necessidades, tecnologias e metodologias para o enriquecimento semântico de dados publicados na *web*. Além disso, pretende-se ampliar a busca realizada para o estado da arte da temática estabelecida para dez anos em mais bases de dados de pesquisas científicas.

Como trabalho futuro, pretende-se, também, ampliar essa pesquisa do estado da arte para outras temáticas relacionadas, como as taxonomias em sistemas educacionais e sua influência no desenvolvimento da competência informacional, propondo um modelo taxonômico para isso. Para o enriquecimento semântico de objetos, pretende-se avaliar se de fato o uso de taxonomias e ontologias pode contribuir para o seu objetivo e ainda propor um novo modelo para esse enriquecimento semântico de dados no contexto corporativo para publicação de dados em *linked data*.

Notas

- [1] ANSI/NISO Z39.19-2005: **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda: NISO, 172 p. 2005. Disponível em: <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>. Acesso em: 28 mar. 2023.
- [2] Repositório Institucional da Universidade Federal de Minas Gerais (RI/UFMG) é um sistema que possibilita o acesso à base de dados de produções científicas da Universidade Federal de Minas Gerais. Disponível em: <https://repositorio.ufmg.br>. Acesso em: 22 mar. 2023.
- [3] *Springer Link* é uma editora mundial baseada na publicação de livros texto, livros de referência acadêmica, e periódicos de artigos, isto é, uma base de dados de estudos científicos. Disponível em: <https://link.springer.com>. Acesso em: 22 mar. 2023.
- [4] *Scientific Electronic Library Online* (SciELO) é uma biblioteca digital de livre acesso e modelo. Disponível em: <https://www.scielo.org/>. Acesso em: 22 mar. 2023.
- [5] *Wiley Online Library* concentra suas atividades em diferentes áreas da ciência, publica material para a comunidade científica, para pesquisadores, práticos (técnicos) e estudantes. Disponível em: <https://www.wiley.com/en-us>. Acesso em: 22 mar. 2023.
- [6] *Linked Data* são princípios e boas práticas para publicação de dados na *web*. Estes princípios denominados *Linked Data* têm como objetivo fundamental facilitar a integração de dados de diferentes fontes, de forma a torná-los compreensíveis também para as máquinas (BIZER; HEATH; BERNERS-LEE, 2009).
- [7] *Europeana Data Model* (EDM) é um modelo para dados proposto pela European - uma biblioteca virtual com a maior coleção *online* de arte. Sendo considerado a *Europeana Data Model primer* (2013) como referência. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf. Acesso em: 27 mar. 2023.

[8] *Bibliographic Framework Initiative* (BIBFRAME) é um modelo de metadados que foi desenvolvido para publicar os registros bibliográficos na *web*, utilizando os princípios do *Linked Data*. Disponível em: <http://www.loc.gov/bibframe/>. Acesso em: 27 mar. 2023.

Referências

ANDERSON Elizabeth J.; MIDGLEY Katherine J.; HOLCOMB Phillip J.; RIÉS, Stephanie K. Taxonomic and thematic semantic relationships in picture naming as revealed by Laplacian-transformed event-related potentials. *Psychophysiology*, [S.l.], v. 59, n. 11, Nov. 2022.

AGANETTE, Elisângela Cristina. **Taxonomias corporativas**: um estudo sobre definições e etapas de construção fundamentado na literatura publicada. 2010. 104f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <https://repositorio.ufmg.br/handle/1843/ECID-87BJSR>. Acesso em: 19 jan. 2023.

AQUINO, Idalécio; CARLAN, Eliana; BRASCHER, Marisa. Princípios classificatórios para a construção de taxonomias. **PontodeAcesso**, Salvador, v. 3, n. 3, p. 196-215, dez. 2009. Disponível em: <https://www.periodicos.ufba.br/index.php/revistaici/article/view/3626/2744>. Acesso em: 20 jan. 2023.

AUSTIN, Patrik. Theory of language: a taxonomy. *SN Soc Sci*, [S.l.], v. 1, n. 78, p. 1-24, 2021. Disponível em: <https://doi.org/10.1007/s43545-021-00085-x>. Acesso em: 26 mar. 2023.

BARBOSA, Alice Príncipe. Classificações facetadas. **Ciência da Informação**, Brasília, DF, v. 1, n. 2, p. 73-81, jan. 1972. Disponível em: <http://revista.ibict.br/ciinf/article/view/10/10>. Acesso em: 07 fev. 2023.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked Data: The story so far. **Semantic services, interoperability and web applications**: emerging concepts. United States of America: 2009. p. 205-227. Disponível em: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far. Acesso em: 27 mar. 2023.

BRASCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRASCHER, M. (org.). **Passeios pelos bosques da informação**: estudos sobre representação e organização da informação e do conhecimento. Brasília: IBICT, 2010. p. 147-176.

DIAS, Célia da Consolação; BARROSO, Juliana Martins de Castro; ROSADO, Keila Mara Lara; SANTOS, Sarah Rúbia de Oliveira. SOC e Interoperabilidade: desafios e tendências no âmbito da Conferência Internacional da ISKO. **FRC: Fronteiras da Representação do Conhecimento**, Belo Horizonte, v. 1, n. 2, p. 1-26, 2022.

EMYGDIO, Jeanne Louize. **Interoperabilidade semântica orientada por ontologia para a Ciência da Informação**: a metodologia Onto4All-Interoperability como resultado de estudo de caso no domínio de energia. 2021. 325f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2021. Disponível em:

<https://repositorio.ufmg.br/handle/1843/42050>. Acesso em: 19 fev. 2023.

FAGUNDES, Priscila Basto; FREUND, Gislaiane Parra; VITAL, Luciane Paula; BARROS, Camila Monteiro de; MACEDO, Douglas Dyllon Jeronimo de. Taxonomias, ontologias e tesaurus: possibilidades de contribuição para o processo de Engenharia de Requisitos. **Em Questão**, Porto Alegre, v. 26, n. 1, p. 237-254, jan. 2020.

FIRMINO, Geraldo José Rodrigues. **Reúso de recursos da web semântica para a construção de vocabulários controlados no contexto da ciência da informação**. 2019. 175f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <https://repositorio.ufmg.br/handle/1843/32372>. Acesso em: 19 fev. 2023.

HJØRLAND, Birger. Theory and metatheory of information science: a new interpretation. *Journal of Documentation*, [S.l.], v. 54, n. 5, p. 606-621, dez. 1998. Disponível em: <http://www.emeraldinsight.com/doi/abs/10.1108/EUM0000000007183>. Acesso em: 19 mar. 2023.

HJØRLAND, Birger. Concept theory. *Journal of the American Society for Information Science and Technology*, [S.l.], v. 60, n. 8, p. 1519-1536, 2009. Disponível em: <https://dl.acm.org/doi/abs/10.5555/1572678.1572685>. Acesso em: 22 mar. 2023.

KITCHENHAM, Barbara; CHARTERS, Stuart. **Guidelines for performing systematic literature reviews in software engineering**. Technical Report EBSE-2007-01. School of Computer Science and Mathematics: Keele University. 2007. Disponível em: https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf. Acesso em: 18 fev. 2023.

LIMA, Flávia Danielle Santiago; GOMES NETO, José Mário Wanderley. Autocontenção à brasileira? Uma taxonomia dos argumentos jurídicos (e estratégias políticas?) explicativo(a)s do comportamento do STF nas relações com os poderes majoritários. **Revista de Investigações Constitucionais**, Curitiba, v. 5, n. 1, p. 221-247, jan./abr. 2018.

LIRA, Márcio Ângelo Bezerra de. **Uma abordagem para enriquecimento semântico de metadados para publicação de dados abertos**. 2014. 94f. Dissertação (Mestrado em Ciência da Computação) - Centro de Informática, Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <https://repositorio.ufpe.br/handle/123456789/11570>. Acesso em: 24 mar. 2023.

LISKA, Geraldo José Rodrigues. **O estudo do léxico na sala de aula [manuscrito]**: investigação do ensino dos processos semânticos de formação de palavras sob a perspectiva da Semântica de Contextos e Cenários (SCC). 2018. 265f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <http://hdl.handle.net/1843/LETR-AYKMP9>. Acesso em: 19 fev. 2023.

- LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. **Data on the Web Best Practices**. [online]. W3C, 2017. Disponível em: <https://www.w3.org/TR/dwbp>. Acesso em: 27 mar. 2023.
- LOPES, Pâmela; AGANETTE, Elisângela; MACULAN, Benildes. Taxonomia corporativa e taxonomia facetada: usos e aplicações na ciência da informação no Brasil. **Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información**, [S.l.], v. 34, n. 82, p. 159-173. 2020. DOI <http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58062>. Acesso em: 22 mar. 2023.
- MAIA, Lucinéia Souza. **Extração e explicitação de relações semânticas para a representação do conhecimento de documentos acadêmicos: um estudo de caso a partir de uma estrutura classificatória**. 2018. 248f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/VAFA-BB5JHW>. Acesso em: 19 fev. 2023.
- MACULAN, Benildes Coura Moreira dos Santos; LIMA, Gercina Ângela 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. **Anais [...]**. Brasília: ANCIB, 2011. p. 696-714. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/174991>. Acesso em: 20 fev. 2023.
- MAIA, Lucinéia Souza; LIMA, Gercina Ângela Borém de Oliveira; MACULAN, Benildes Coura Moreira dos Santos. Taxonomia dos tipos de relações semânticas para a organização e a representação do conhecimento: uma proposta a partir da literatura. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 10, n. 1, p. 1-22, jan./jun. 2017. Disponível em: <https://brapci.inf.br/index.php/res/v/104795>. Acesso em: 21 jan. 2023.
- MENDES, Paula Raphisa. **Taxonomia Corporativa Facetada no Apoio ao Processo de Avaliação para Autorização de Cursos de Instituições de Ensino Superior pelo INEP**. 2022. 99f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2022. Disponível em: <http://hdl.handle.net/1843/48840>. Acesso em: 20 mar. 2023.
- NAZÁRIO, Claudiane Emanuele. **Proposta de metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do Linked Data**. 2018. 140f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/ECIP-B3UJKW>. Acesso em: 19 fev. 2023.
- NAZÁRIO, Claudiane Emanuele; DIAS, Célia da Consolação. Proposta para avaliação das técnicas e dos recursos para o enriquecimento semântico de objetos publicados em Linked Data. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 11., 2018. Londrina. **Anais [...]**. Londrina: ANCIB, 2018.
- OLIVEIRA, Elaine Diamantino. **Proposta de um modelo de hipertexto com abordagem semântica para a representação do conhecimento no domínio temático da Intensificação Agropecuária**. 2018. 227f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUOS-B2JP2A>. Acesso em: 20 mar. 2023.
- PETERSEN, Kai; FELDT, Robert; MUJTABA, Shahid; MATTSSON, Michael. Systematic mapping studies in software engineering. In: INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING, 12., 2007, Swindon. **Proceedings [...]**. Swindon: BCS, 2007. p. 68-77. Disponível em: <https://dl.acm.org/doi/10.5555/2227115.2227123>. Acesso em: 10 jan. 2023.
- PETTICREW, Mark; ROBERTS, Helen. **Systematic Reviews in the Social Sciences: A Practical Guide**, Blackwell Publishing, 2005. ISBN 1405121106. Disponível em: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470754887>. Acesso em: 18 fev. 2023.
- RANGANATHAN, Shivali Ramamrita. **Colon classification, basic classification**. 6. ed. New York: Bombay, 1960.
- RANGANATHAN, Shiyali Ramamrita. **Prolegomena to Library Classification**. Bom- bay: A. P. House, Ed. 1967.
- SANTOS, Thais Helen do Nascimento. A taxonomia e a folksonomia na representação da informação de fotografias. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 23, n. 1, p. 89-10, 2018. Disponível em: <https://doi.org/10.1590/1981-5344/2395>. Acesso em: 10 mar. 2023.
- SORRENTINO, Serena; BERGAMASCHI, Sonia; FUSARI, Elisa; BENEVENTANO, Domenico. Semantic Annotation and Publication of Linked Open Data. In: INTERNATIONAL CONFERENCE – ICCSA, 13., 2013, **Proceedings [...]**. Ho Chi Minh City, Vietnam, 2013. p. 462-474. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-39640-3_34. Acesso em 25 mar. 2023.
- THAKKER, Dhavalkumar; DESPOTAKIS, Dimoklis; DIMITROVA, Vania; LAU, Lydia; BRNA, Paul. Taming Digital Traces for Informal Learning: A Semantic-driven Approach. In: EUROPEAN CONFERENCE ON TECHNOLOGY ENHANCED LEARNING, 7., 2012, Berlin. **Proceedings [...]** Heidelberg: Springer-Verlag, 2012. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-33263-0_27 Acesso em: 26 mar. 2023.
- VIANA, Jordânia Quintão. **Metodologia para a construção de taxonomia corporativa facetada**. 2022. 168f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2022. Disponível em: <https://repositorio.ufmg.br/handle/1843/44638>. Acesso em: 19 fev. 2023.