

**Organização do conhecimento e gestão dos tesouros na web
semântica**

***Organization of knowledge and management of thesauris on the
semantic web***

***Organización del conocimiento y gestión de tesauros em la web
semântica***

***Organisation des connaissances et gestion des thésaurus sur le
web sémantique***

**Adriana Carla Ribeiro dos Santos¹
Thiago Henrique Bragato Barros²
Rita do Carmo Ferreira Laipelt³**

Autor principal para Correspondência

Adriana Carla Ribeiro dos Santos

E-mail: s.adrianacarla@yahoo.com.br

ORCID: <https://orcid.org/0000-0003-0885-1478>



Submetido em: 24/09/2021

Aceito em: 06/10/2021

Publicado em: 29/12/2021

¹ Mestranda no Programa de Pós-Graduação em Ciência da Informação pela Universidade Federal do Pará (PPGCI/UFPA).

² Professor Adjunto no departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS), professor permanente nos Programas de Pós-graduação em Ciência da Informação da UFRGS e UFPA.

³ Professora Adjunta do Departamento de Ciências da Informação da faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul (UFRGS).

RESUMO

Introdução: O tesauro é um importante instrumento para representar e recuperar a informação em áreas especializadas. Sua relevância configura-se no controle terminológico dos documentos. A Organização do Conhecimento é um campo científico interdisciplinar que vem buscando, ao longo do tempo, consolidar sua identidade. Hoje, os sistemas de busca são os principais meios de localização de informação na *web*. Porém, há uma certa confiabilidade nesses resultados de busca, o que significa que há uma insatisfação com a mesma. **Objetivo:** Visa-se proporcionar uma análise literária sobre tesauro nos contextos da Organização do Conhecimento e da *Web Semântica*, por meio da bibliometria e análise de conteúdo. **Metodologia:** Revisão de literatura na área especializada, além de pesquisa na base de dados da BRAPCI, a fim de investigar a produção bibliográfica sobre o tema tesauro no Brasil entre 2015 a 2021. **Resultados:** A principal proposição é uma breve sistematização do tesauro enquanto Sistema de Organização do Conhecimento, como forma de relacionar e representar o instrumento em um domínio específico. **Conclusões:** Ressalta-se a importância de estudos sobre as bases teóricas e epistemológicas para oferecer um percurso teórico explícito que contribua na elaboração de uma metodologia de gestão dos tesouros para o profissional da informação. Foi possível identificar que os estudos da Organização do Conhecimento no contexto da Ciência da Informação vêm ganhando dimensões proporcionais devido à aceleração do uso de tecnologias da *web*.

Palavras-Chave: Tesauro; *Web Semântica*; Sistema de Organização do Conhecimento.

ABSTRACT

Introduction: The thesaurus is an important tool to represent and retrieve information in specialized areas. Its relevance is configured in the terminological control of documents. Knowledge Organization is an interdisciplinary scientific field that has been seeking, over time, to consolidate its identity. Today, search engines are the main means of finding information on the web. However, there is a certain unreliability in these search results, which means that there is dissatisfaction with the search. **Objective:** The aim is to provide a literary analysis on thesaurus in the contexts of Knowledge Organization and the Semantic Web, through bibliometrics and content analysis. **Methodology:** Literature review in the specialized area, in addition to research in the BRAPCI database, in order to investigate the bibliographic production on the thesaurus in Brazil,

between 2015 and 2021. Results: The main proposition is a brief systematization of the thesaurus as a System of Knowledge Organization, as a way to relate and represent the instrument in a specific domain. Conclusions: We emphasize the importance of studies on the theoretical and epistemological bases to offer an explicit theoretical path that contributes to the development of a thesaurus management methodology for the information professional. It was possible to identify that the studies of Knowledge Organization in the context of Information Science have been gaining proportional dimensions, due to the acceleration of the use of web technologies.

Key words: Thesaurus; Semantic Web; Knowledge Organization System.

RESUMEN

Introducción: El tesauro es una herramienta importante para representar y recuperar información en áreas especializadas. Su relevancia se configura en el control terminológico de los documentos. La Organización del Conocimiento es un campo científico interdisciplinario que ha buscado, a lo largo del tiempo, consolidar su identidad. Hoy en día, los motores de búsqueda son el principal medio para encontrar información en la web. Sin embargo, existe una cierta falta de fiabilidad en estos resultados de búsqueda, lo que significa que hay insatisfacción con la búsqueda. Objetivo: El objetivo es proporcionar un análisis literario sobre tesauros en los contextos de Organización del Conocimiento y Web Semántica, a través de bibliometría y análisis de contenido. Metodología: Revisión de la literatura en el área especializada, además de la investigación en la base de datos BRAPCI, para investigar la producción bibliográfica sobre el tema del tesauro en Brasil, entre 2015 y 2021. Resultados: La propuesta principal es una breve sistematización del tesauro como un Sistema de Organización del Conocimiento, como forma de relacionar y representar el instrumento en un dominio específico. Conclusiones: Destacamos la importancia de los estudios sobre las bases teóricas y epistemológicas para ofrecer un camino teórico explícito que contribuya al desarrollo de una metodología de gestión de tesauros para el profesional de la información. Se pudo identificar que los estudios de Organización del Conocimiento en el contexto de las Ciencias de la Información han ido ganando dimensiones proporcionales, debido a la aceleración del uso de tecnologías web.

Palabras clave: Tesauro; Web semántica; Sistema de organización del conocimiento.

RÉSUMÉ

Introduction: L'organisation des connaissances est un domaine scientifique interdisciplinaire qui a cherché, au fil du temps, à consolider son identité. Aujourd'hui, les systèmes de recherche constituent le principal moyen de localiser des informations sur le web. Toutefois, ces résultats de recherche présentent un certain manque de fiabilité, ce qui signifie qu'ils suscitent une certaine insatisfaction. Le Web sémantique a pour principe de nous apporter des informations diverses afin que l'utilisateur et les machines soient en mesure de comprendre et d'interpréter les données. **Objectifs:** Cet article vise à fournir une analyse littéraire sur la construction et la maintenance des thésaurus dans les contextes de l'organisation des connaissances et du Web sémantique, afin d'offrir une voie théorique qui contribue à l'élaboration d'une méthodologie de gestion des thésaurus pour les professionnels de l'information. **Méthodologie:** une recherche descriptive et exploratoire, avec revue bibliographique. L'article est organisé selon les sections suivantes : tout d'abord, une discussion théorique de l'organisation des connaissances et des systèmes d'organisation des connaissances est présentée. Ensuite, une approche de la sémantique Web et de la théorie fondamentale de la construction et de la gestion des systèmes d'information réglementés par la norme (ISO 25964, 2011). **Conclusions:** Il a été possible d'identifier que les études de l'organisation de la connaissance dans le contexte des sciences de l'information ont gagné des dimensions proportionnelles, en raison de l'accélération de l'utilisation des technologies web. Les chercheurs dans le domaine de l'organisation des connaissances travaillent à la recherche de nouvelles théories et méthodologies qui contribuent au référentiel théorique des systèmes d'organisation des connaissances.

Mots clés: Organisation de la connaissance. Systèmes d'organisation des connaissances. Web sémantique. Thésaurus.

1 INTRODUÇÃO

A Organização do Conhecimento (OC) é um ramo de especialidade da Ciência da Informação que tem como base o estudo das teorias, metodologias, instrumentos e produtos para o acesso ao conhecimento (HJØRLAND, 2007). Consiste em um campo científico interdisciplinar que vem buscando, ao longo do tempo, consolidar sua identidade, como é

apresentado na literatura da *International Society of Knowledge Organization* (ISKO).

No contexto da Organização do Conhecimento, os Sistemas de Organização do Conhecimento (SOC), também conhecidos por (KOS), um acrônimo do inglês *Knowlegde Organization System*, são considerados sistemas conceituais semanticamente organizados que analisam os termos, as definições, os relacionamentos e as propriedades dos conceitos.

Na representação do conhecimento, SOC são instrumentos usados para a organização e recuperação da informação de bibliotecas, museus e arquivos, e seu objetivo é a padronização terminológica que facilita e orienta a indexação e os usuários. Em sua estrutura, os SOC variam de um modelo simples até o multidimensional; entretanto, suas funções incluem controle de sinônimos ou equivalentes, ambiguidade e relacionamentos semânticos entre conceitos.

Os Tesouros, nosso objeto de estudo, são considerados vocabulários controlados e estruturados que apresentam relações hierárquicas, associativas e de equivalências entre termos e conceitos. São amplamente difundidos na Biblioteconomia, mas pouco explorados na Arquivologia pelo fato de que muitos profissionais arquivistas desconhecem sua operabilidade.

A *Web Semântica* tem contribuído de forma significativa para a renovação dos tesouros como suporte para buscas semânticas e outros serviços. Hoje, as ferramentas que gerenciam tesouros permitem que sejam criados, editados e consultados.

No que diz respeito à normativa que rege a gestão dos tesouros, fica clara a importância da sua construção e manutenção, principalmente pelo fato de que ela é realizada por organismos internacionais e nacionais, pois implementa sua regulamentação nas propostas de

necessidades de controle de vocabulário em contexto digital, bem como no uso de tecnologias avançadas para o compartilhamento de conceitos e termos.

Esta pesquisa foi desenvolvida por meio de revisão de literatura sobre tesauro, observando sua fundamentação conceitual e estrutural de elaboração e construção enquanto sistema de organização do conhecimento. O objetivo foi identificar os princípios teóricos-metodológicos encontrados na literatura da área e os princípios teóricos selecionados nas publicações sobre o tema, por meio de análise bibliométrica e de conteúdo.

2 REFERENCIAL TEÓRICO

Considerando o tesauro a temática central desta pesquisa, o referencial teórico trata, inicialmente, das bases teórico-metodológicas de seu desenvolvimento. Posteriormente, aborda o tesauro enquanto sistema de organização do conhecimento e, por último, analisa-o no contexto da *web* semântica.

2.1 As bases teórico-metodológicas do desenvolvimento de Tesauro

O tesauro é uma ferramenta de representação da informação, caracterizado em determinado domínio com diferentes tipos de níveis de terminologia e padronização, o que auxilia na indexação (BARITÉ, 2008).

Além do mais, consiste em um tipo de vocabulário controlado, capaz de estabelecer a ligação entre o usuário e o conteúdo temático de um acervo, muito usado na Biblioteconomia, quase nulo na Arquivologia. Portanto, os tesauros são usados como instrumentos de organização do conhecimento e recuperação da informação.

No *Diccionario de Organización del Conocimiento: clasificación, indización, terminología*, Barité (2015) mostra uma relação de características para o termo tesouro:

1. Tipo de sistema de organización del conocimiento que se integra con términos analizados y normalizados que guardan entre sí relaciones semánticas y funcionales. El tesouro se organiza bajo fuerte control terminológico, con objeto de proporcionar un instrumento idóneo para el almacenamiento y la recuperación de la información en áreas especializadas. Puede ser monolingüe, monolingüe con equivalencias o multilingüe, conforme a la cobertura idiomática que proponga. En ciertos casos, agrega una notación. El tesouro también es llamado tesoro. // 2. Repertorio que inventaría, con la aspiración de exhaustividad, el conjunto de unidades léxicas de una lengua. Se diferencia del diccionario general de una lengua en que recopila diacrónicamente y sin criterios de selectividad (BARITÉ, 2015, p.156).

Segundo Barité, tesouro é caracterizado como um “tipo de linguagem documentária composta de termos analisados e normalizados que mantêm entre si relações semânticas e funcionais”, ou seja, o tesouro “é organizado sob rigoroso controle terminológico, com o objetivo de fornecer um instrumento idôneo para o armazenamento e recuperação de informações em áreas especializadas”. Afirma-se, ainda, que, para facilitar o acesso ao conhecimento, pode ser apresentado em uma ou várias línguas, isto é, monolíngue, monolíngue com equivalências, ou multilíngue, de acordo à língua de cobertura estabelecida na construção do tesouro. Por fim, destaca-se que o tesouro, em certos casos, agrega notação (BARITÉ, 2015).

Existem diversas definições a respeito do tesouro e, por essa razão, utilizamos um quadro-resumo (Quadro 1) de algumas definições de tesouro segundo normas e autores:

Quadro 1 - Definições de tesouro

TESAURO	
“[...] vocabulário controlado e dinâmico abrangendo área específica do conhecimento. Em sua estrutura, patenteia as relações vigentes entre os	(IBICT, 1984, p. 5).

termos ou descritores – sinonímicas hierárquicas e outras – que, no conjunto, constitui a linguagem de indexação”.	
“[...] vocabulário de uma linguagem de indexação controlado e organizado formalmente com objetivo de explicitar as relações a priori entre conceitos (por exemplo, mais genérico que... ou mais específico que...)”. Para isso, descreve a linguagem de indexação como: “conjunto controlado de termos extraídos da linguagem natural e utilizados para representar de forma breve os assuntos dos documentos”.	(ISO 2788, 1986).
Tesouro é um vocabulário de termos relacionados genérica e semanticamente sobre determinada área de conhecimento.	(MOTTA, 1987)
“[...] Linguagem documentária dinâmica que contém termos relacionados semântica e logicamente, cobrindo de modo compreensivo um domínio do conhecimento”.	(GOMES, 1990, p. 16).
“[...] lista estruturada de conceitos destinados a representar de maneira unívoca o conteúdo dos documentos e das consultas dentro de um sistema documental determinado [...] inclui descritores, não-descritores, relações hierárquicas e de associação e equivalências linguísticas”.	(VAN SLYPE, 1991, p.23-24).
“[...] linguagem documentária, construída por meio de unidades conceituais, extraídas da linguagem formal de uma área específica do conhecimento científico ou técnico. Sua estrutura sugere a ideia de sistema, visto que os conceitos relacionam-se entre si e são representados por termos. Cada termo, por sua vez, possui vinculação com outro termo, por meio de relação de equivalência, de hierarquia ou de associação. O tesouro, utilizado para a organização e recuperação da informação, constitui-se em importante “[...] instrumento de apoio às pesquisas científicas nas áreas de conhecimento”.	(FUJITA, 1992, p. 23-24).
“[...] linguagem documentária que representa de forma normalizada os conceitos de uma área específica através de termos que se manifestam em estruturas lógico-semânticas”.	(TÁLAMO; LARA; KOBASHI, 1992).
“[...] vocabulário controlado de uma linguagem de indexação, formalmente	(UNESCO, 1993, p. 14).

<p>organizado para explicitar as relações a priori entre conceitos (por exemplo, como genéricas e específicas).</p>	
<p>“[...] vocabulário especializado, normalizado, pós-coordenado, usado com fins documentários, onde os elementos linguísticos que o compõem, termos simples ou compostos, encontram-se relacionados entre si sintática e semanticamente”.</p>	<p>(CURRÁS, 1998).</p>
<p>“[...] sistema de classificação temática ou facetada, cuja estrutura básica está conformada por uma relação de descritores que representam ou descrevem autoridades ou conteúdos temáticos. [...] por meio de unidades linguísticas, semânticas e suas relações, extraídas da linguagem formal de uma disciplina ou área específica do conhecimento que [...] se torna um instrumento de representação e recuperação da informação”.</p>	<p>(NAUMIS PEÑA, 2000).</p>
<p>“[...] instrumento apropriado para transmitir conceitos e as relações recíprocas destes, semelhantemente ao que ocorre com os termos expressos na linguagem dos documentos”.</p>	<p>(DODEBEI, 2002, p. 67).</p>
<p>“[...] como função – “um instrumento de controle terminológico que permite traduzir a linguagem natural dos documentos, dos indexadores e dos usuários, numa ‘linguagem sistêmica’ mais rígida (linguagem documentária, linguagem do sistema de informação)”. Como estrutura – “um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente, que cobre um campo específico de conhecimentos”. (ROBREDO, 2005, p. 157-158). “[...] vocabulário controlado organizado em uma ordem conhecida e estruturada de modo que os vários relacionamentos entre os termos sejam identificados e indicados claramente por meio de orientações normativas”.</p>	<p>(ANSI/NISO-Z39.19, 2005, p. 9).</p>
<p>“[...] linguagens de estruturas combinatórias e pós-coordenadas, constituídas de termos – unidades linguísticas provenientes da linguagem de especialidade e da linguagem natural –, denominados de descritores, providos de relações sintático-semânticas, referentes a domínios científicos especializados, possibilitando a</p>	<p>(BOCCATO; RAMALHO; FUJITA, 2008, p. 201).</p>

representação temática do conteúdo de um documento, bem como a recuperação da informação”.	
Tipo de sistema de organização do conhecimento integrado aos termos analisados e padronizados que mantêm relações semânticas e funcionais entre si. O dicionário de sinônimos é organizado sob forte controle terminológico, a fim de fornecer um instrumento ideal para armazenar e recuperar informações em áreas especializadas. [...]	(BARITÉ, 2015).

Fonte: (Elaborado pelos autores, 2021)

De acordo com as definições referidas no Quadro 1, podemos destacar que o tesauro pode ser considerado um importante instrumento e o mais adequado para representar e recuperar a informação em áreas especializadas. Sua relevância configura-se no controle terminológico dos documentos de um sistema de informação especializada e no emprego dos mesmos termos para representar os documentos compostos pelos conceitos, quando submetido à busca de um assunto. Lembrando que o propósito da Lei de Acesso à Informação (LAI) é garantir o acesso público aos documentos de arquivo, permitindo seu acesso e recuperação.

Observa-se, ainda, no contexto da recuperação da informação, que os planos de classificação falham no sentido de explicitar as relações entre os termos, além de não apresentarem em sua estrutura categorias temáticas.

O tesauro, ao contrário, oferece subsídios para o relacionamento entre os termos (hierarquia, associativa e de equivalência) que possibilitam uma forma mais detalhada da informação. As cadeias conceptuais elaboram as árvores semânticas que partem dos termos mais gerais aos específicos.

A UNESCO apresentou como norma uma lista de abreviaturas e símbolos que são utilizados em tesouros como prefixos dos termos. As

abreviaturas e símbolos apresentados em português têm uma explicação sobre seu significado (assim como de seus equivalentes em inglês) e são usados para indicar a relação ou função do termo ou nota que se apresenta, conforme destacado no quadro 2:

Quadro 2 - Lista de termos

Português	Inglês
TG = Termo Genérico. O termo que segue refere-se a um conceito com conotação mais ampla. Superordenado.	BT = <i>Broader Term</i>
TGM = Termo Genérico Maior. O termo que segue é o nome da classe mais ampla à qual pertence o conceito específico, também usado, às vezes, na seção alfabética de um tesouro.	BTG = <i>Broader Term (Generic)</i>
TGP = Termo Genérico Partitivo. O termo que segue representa o todo em relação à parte.	BTP = <i>Broader Term (Partitive)</i>
TE = Termo Específico. O termo que segue refere-se a um conceito com conotação mais específica. Subordinado.	NT = <i>Narrower Term</i>
TEP = Termo Específico Partitivo. O termo que segue representa a parte em relação ao todo.	NTP = <i>Narrower Term (Partitive)</i>
TR = Termo Relacionado. O termo que segue está associado, mas não é sinônimo nem termo genérico ou termo específico.	RT = <i>Related Term</i>
NE = Nota Explicativa (ou Nota de Escopo). Nota que se junta a um termo para indicar seu significado específico dentro de uma linguagem de indexação.	SN = <i>Scope Note</i>
UP = Usado Para. O termo que segue é um sinônimo ou um quase sinônimo do termo preferido.	UF = <i>Used For</i>
USE. O termo que segue é o termo preferido quando se deve escolher entre sinônimos ou quase sinônimos.	USE

Fonte: (UNESCO, 1993).

As abreviaturas listadas no Quadro 2 são consideradas convenções reconhecidas que aparecem em diversos tesouros publicados e possuem valor *mnemônico*, embora se reconheça que também sejam dependentes do idioma (UNESCO, 1993, p. 79). De modo geral, podemos dizer que as relações básicas de um tesouro – relações de equivalência, relação

hierárquica e relação associativa – configuram-se por meio das abreviaturas e símbolos da seguinte forma (Quadro 3):

Quadro 3 - Relações básicas de um tesouro

RELAÇÕES DE EQUIVALÊNCIA	RELAÇÃO HIERÁRQUICA	RELAÇÃO ASSOCIATIVA
USE = precede o termo preferido.	TGM = Termo genérico Maior de uma hierarquia	TR = Termo Relacionado
UP = precede termo não preferido.	TG = Termo Genérico	
	TGP = Termo Genérico Partitivo	
	TE = Termo Específico	
	TEG = Termo Específico Genérico	
	TEP = Termo Específico Partitivo	

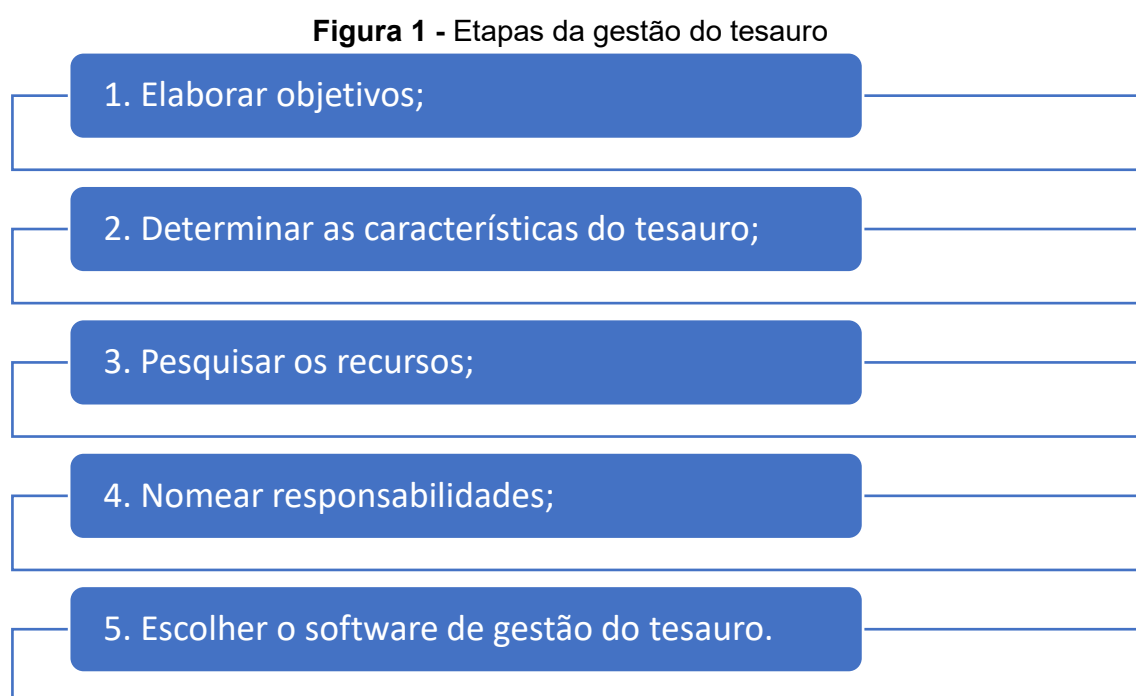
Fonte: (UNESCO, 1993)

O tesouro, enquanto linguagem documentária, elaborado com interface na terminologia, possibilita uma representação bem mais adequada e compatível com a linguagem dos usuários, permitindo, assim, uma recuperação da informação precisa e pertinente. Em outras palavras, o tesouro serve como índice para o plano de classificação, de forma que ambos se complementam, visto que oferecem duas formas distintas, mas não opostas de apresentação dos termos, sendo que os tesouros dispõem de mais recursos, o que evidencia as características e a relação estabelecida entre termos (SMIT; KOBASSHI, 2003, p.40).

A norma ISO 25964, publicada em 2011, estabelece a gestão dos tesouros, trata da sua construção e manutenção e é de grande importância, uma vez que é elaborada por organismos internacionais e nacionais e é responsável por implementar toda sua regulamentação.

A normatização tem demonstrado avanços significativos no que diz respeito às necessidades de controle de vocabulário no âmbito da *web*, aplicando tecnologias avançadas e garantindo a gestão e compartilhamento de conceitos e termos.

A norma diz que a gestão do tesouro requer um planejamento para que os interesses e objetivos sejam reunidos e aplicados em sua elaboração. A todo planejamento são aplicadas diretrizes e a gestão segue as seguintes etapas, conforme Figura 1:



Fonte: Elaborado pelos autores (2021) com base na ISO (2011).

Os principais pontos a serem considerados nos objetivos do planejamento estabelecido pela ISO são: “O tesouro deve ser usado para quê, e por quem; ele será limitado pelas restrições do *software* existente com o qual deverá ser usado; quão conhecedores serão os usuários da área de assunto do tesouro e de uso do tesouro” (ISO,2011).

De acordo com a normatização, o planejamento, sem dúvida, é de fundamental importância, pois, ao estabelecer esses pontos, tem-se um estudo prévio com diretrizes dos objetivos, assim como uma organização das tarefas predefinidas para as pessoas envolvidas na construção do tesouro.

A próxima etapa determinada pela norma (ISO, 2011) refere-se às características do tesouro, que devem ser delimitadas da seguinte forma:

- I. Qual o formato do tesauro? Será impresso, eletrônico ou ambos?
- II. Qual o modelo de apresentação requerido?
- III. Há necessidade de formatos especiais, tais como sistemas de indexação ou busca?
- IV. Quais formatos são previstos para atualizações e quão frequentemente elas serão requeridas?

As definições são essenciais das características, tendo-se em vista o formato de apresentação do tesauro que melhor se ajuste aos objetivos iniciais propostos. Ressalta-se também a possibilidade de se incluir características opcionais, como o armazenamento de informações sobre definições de termos, relações customizadas e presença dos rótulos de nó.

Os formatos previstos para atualizações estabelecidos na (ISO, 2011) são: termos utilizados no plural ou singular, grafia, maiusculização, extensão máxima do termo, caracteres especiais, caracteres especiais e, se houver, número limite de níveis da hierarquia

Quanto aos recursos do tesauro, a norma destaca que “[...] os recursos-chave, tais como as pessoas, o financiamento, as ferramentas de *software* e os recursos de vocabulário, devem ser determinados” (ISO, 2011).

Para o estabelecimento das responsabilidades, cabe ao responsável do projeto a função de delegar as atividades do grupo que participa do projeto.

Por fim, a escolha do *software* de gestão de tesauro representa uma decisão importante, visto que é por meio do suporte oferecido pelo *software* que a eficiência do serviço se manifestará. Antes, no entanto, faz-se necessário avaliar os *softwares* presentes no mercado, uma forma segura de escolher qual se enquadra nos objetivos e orçamento do projeto.

Finalizando as etapas estabelecidas anteriormente, segue o quadro 4 com novas etapas para dar continuidade ao planejamento do tesauro, segundo a (ISO, 2011):

Quadro 4 - Planejamento do Tesauro (continuação)

ORIENTAÇÕES PARA:	ETAPAS
Estágios iniciais da compilação	Coleta e análise dos termos
Construção do tesauro	Elaboração das estruturas hierárquicas; especificidade; relações de equivalência e associativas.
Introdução do tesauro	Configuração do tesauro
Disseminação do tesauro	Integração com a indexação; navegação e pesquisa; formato digital e impresso; diretórios do <i>website</i>
Manutenção do tesauro	Procedimentos de sugestão e de revisão

Fonte: baseado na ISO (2011)

Nos estágios iniciais da compilação, *Early stages of compilation*, a norma recomenda que o tesauro, idealmente, esteja concluído antes do início da compilação da base de dados. Com relação à coleta de termos e conceitos, a ISO destaca que, no estudo dos termos, é necessário identificar as indicações feitas pelos usuários, encontradas nas listas de referências.

Além disso, a norma propõe que o estágio inicial tenha como tarefa: coletar os termos e observar a fonte de cada um, bem como a frequência de ocorrência. No momento da análise dos termos, os mesmos devem ser organizados sistematicamente para que, mais tarde, sejam inseridos no tesauro.

Na construção do tesauro, os termos devem estar organizados em grupos de assuntos/facetas ou hierarquias. Em seguida, deve-se estabelecer os grupos de sinônimos e quase sinônimos, assim como verificar a duplicação de termos, sobreposições ou omissões, verificando-se também o grau de especificidade estabelecido. Ainda, faz-se

necessário definir as estratégias quanto à estrutura hierárquica e quanto ao nível de especificidade. Além disso, a norma trata dos termos agrupados hierarquicamente, em conjunto às relações hierárquicas e de equivalência, e recomenda que as relações associativas sejam incluídas no estágio final.

Para a introdução do tesauro, a norma ISO registra as seguintes informações:

1. A(s) área(s) de assunto(s) coberta(s), com identificação de áreas centrais e marginais;
2. Idiomas nos quais o tesauro é apresentado;
3. Confirmação de qual norma nacional e/ou internacional foi seguida;
4. Noção semântica das convenções, abreviaturas e sinais de pontuação usados de formas não normatizadas;
5. Número total de termos, com totais separados de termos preferidos e não preferidos;
6. Quaisquer regras que regem a seleção de conceitos, incluindo conceitos complexos;
7. Estabelecer regras na seleção de formas dos termos preferidos, incluindo uma referência a qualquer manual de estilo;
8. Seguir e citar uma norma nacional ou internacional apropriada quando possível;
9. Declaração sobre a atualização da política, incluindo a frequência, datas e procedimentos, além do nome e endereço da agência responsável a quem os comentários e sugestões devem ser enviados;
10. Referências das fontes usadas na compilação e revisão do tesauro.

A etapa de divulgação do tesauro descreve a integração com um sistema eletrônico, considerando as necessidades de uso do tesauro,

indexação ou pesquisa. A ISO indica que a última etapa do planejamento do tesauro consiste na manutenção, que deve ser feita durante o período de vida, desde o momento de sua publicação. Trata, ainda, de dois procedimentos importantes: o primeiro refere-se ao mecanismo de mudanças do tesauro, tanto para usuários quanto para indexadores; o segundo, por sua vez, aos procedimentos de revisão, que devem dar prioridade às sugestões e aos termos incluídos regularmente.

2.2 Tesauro como sistema de Organização do Conhecimento

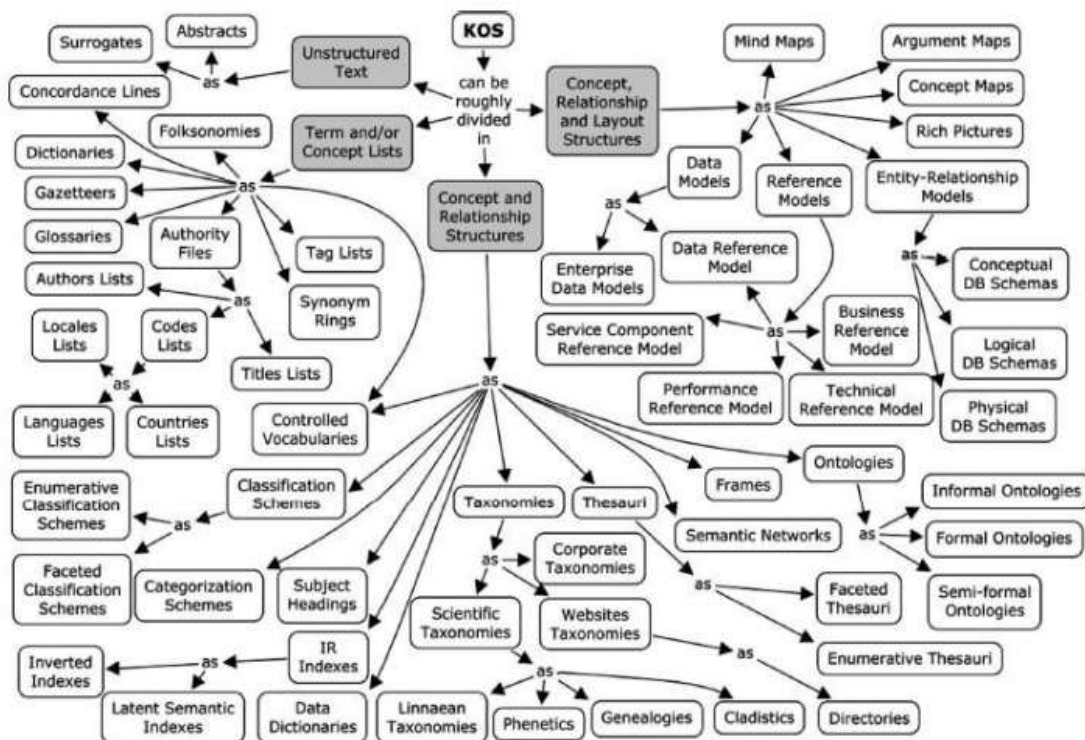
Os Sistemas de Organização do Conhecimento (SOC) são instrumentos de representação do conhecimento que foram estabelecidos pelo *Networked Knowledge Organization Systems Working Group* em uma Conferência da ACM *Digital Libraries*, em 1998, *Pittsburgh* na *Pennsylvania*.

O termo é uma tradução para o português do original inglês “*Knowledge Organization System*” (KOS). A sigla KOS é utilizada com frequência na literatura e, dessa forma, usa-se a sigla SOC em português.

Hjørland (2008) refere-se aos SOC como ferramentas que colaboram na interpretação organizada do conhecimento, conhecidas como ferramentas semânticas.

Segundo Hodge, os (SOC) são todos os tipos de instrumentos que têm como objetivo organizar a informação e promover a ação do conhecimento, incluindo também os sistemas de classificação, que organizam materiais; os cabeçalhos de assunto, que fornecem o acesso mais detalhado; e os catálogos, que controlam versões variantes de informação, como nomes geográficos ou nomes de pessoas e outros esquemas, como as redes semânticas, tesauros, taxonomias e ontologias, conforme mostra a figura 2:

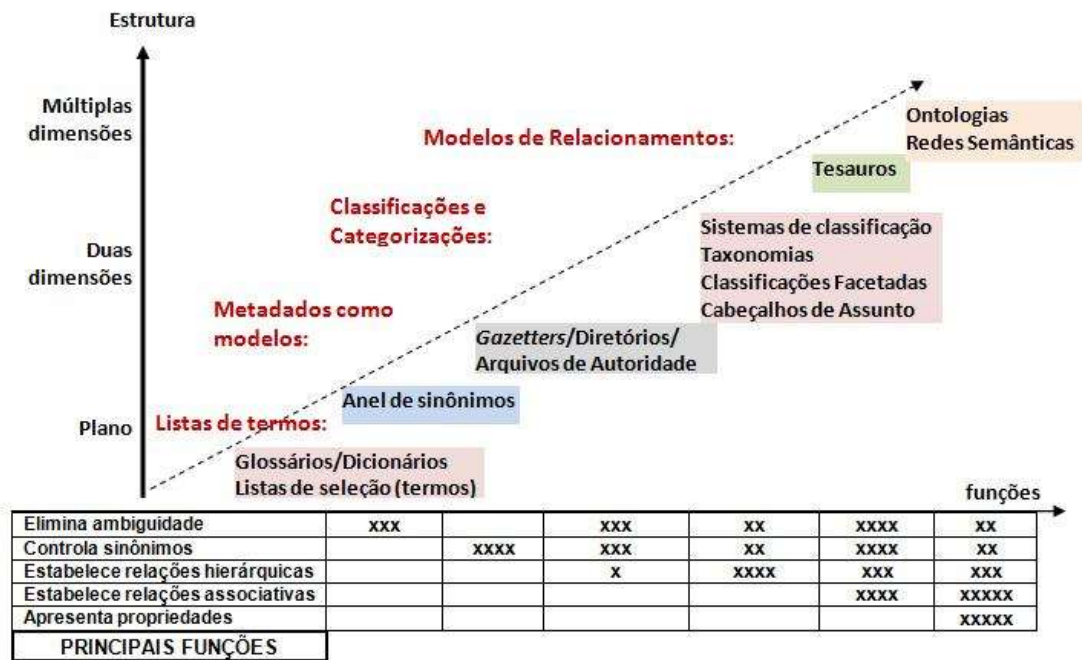
Figura 2 - Os KOS



Fonte: Souza *et al.* (2012)

Já a classificação de Zeng (2008), apresentada na figura 3, mostra dois grupos separados, ou seja, Listas de Termos e Modelos do tipo Metadados, que são diferentes, como se pode observar nas propostas de Hodge e Souza *et al.* (2012).

Figura 3 - Os (SOCs)



Fonte: Zeng (2008)

Ressalta-se, portanto, que os (SOC) são estruturas organizadas que objetivam a construção de padrões abstratos da realidade, representando os conceitos de um domínio. Essas ferramentas semânticas são utilizadas para o tratamento da informação, viabilizando a recuperação da informação em ambiente informatizado ou tradicional.

As classificações e tesouros têm sido utilizados para organizar recursos digitais na Internet. A *Web Semântica* apresenta as ferramentas para desenvolvimento de (SOC), os quais, de fato, estão popularizando-se, principalmente os tradicionais, por causa da necessidade de compartilhamento de padrões orientados por ontologias.

Atualmente, os sistemas de busca são os principais meios de localização de informação na *web*. Porém, há uma certa confiabilidade nesses resultados de busca, o que significa que nem sempre encontramos aquilo que realmente desejamos.

Os tesouros aliados às novas TIC transformam o conteúdo de conhecimento formalizado, identificável e interoperável por meio de

máquinas, considerando a dinamicidade da infraestrutura da *web* (LARA, 2013).

A construção de um tesauro se constitui como uma atividade interdisciplinar, ligada, principalmente, aos campos da Ciência da Computação e da Ciência da Informação.

2.3 O Tesauro no contexto da Web Semântica

A história da *Web Semântica* surgiu em 2001, após a publicação de um artigo da revista *Scientific American* intitulado "*Semantic Web consists of a new format of content for the web that has meaning for computers will start a revolution of new possibilities*", escrito por Tim Berners-Lee, James Hendler e Ora Lassila.

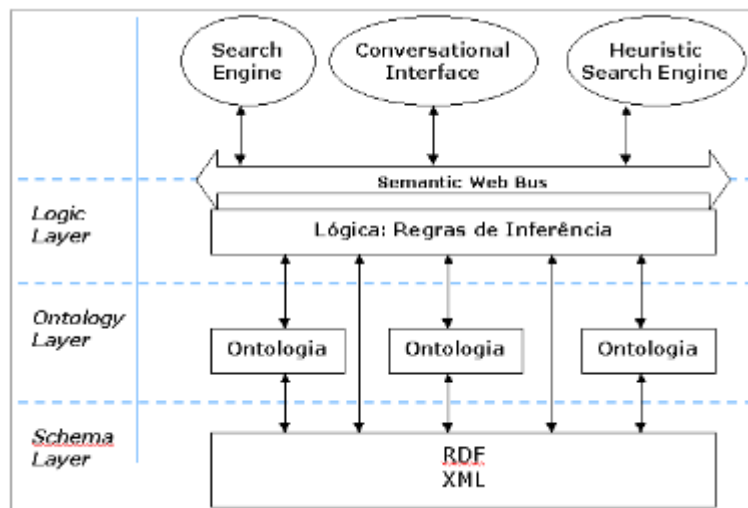
Destaco, aqui, Tim Berners-Lee é britânico, físico, cientista da computação, criador da *World Wide Web*, professor do Instituto de Tecnologia de Massachusetts (MIT) e foi um dos responsáveis pelo surgimento dessa metodologia.

De forma prática, basta-se fazer a busca em um *site* de determinado produto, localizá-lo e, com o sistema disponibilizado, obter o cálculo de frete e verificar o prazo de entrega. Tal procedimento consiste em uma forma prática e segura, além de propícia em tempos de Covid-19, em que milhares de pessoas ficaram impossibilitadas de sair de suas residências, evitando, assim, o contágio da doença.

De modo geral, todas as informações disponíveis na *web* são também tratadas como informações na *Web Semântica*. Isso significa que, se um indivíduo procurasse por um produto, teria como resultado, com esses recursos, não somente os locais disponíveis, mas também valor do frete e o tempo de entrega de forma rápida e ágil.

A Figura 4 apresenta uma arquitetura elaborada por Berners-Lee (2001) para a *Web Semântica*, composta de três camadas:

Figura 4 - *Web Semântica*



Fonte: Elaborado por Berners-Lee (2001)

2.3.1 A camada de esquema (*schema layer*)

A camada esquema tem como função definir os dados do documento e o significado associado desses dados, estruturando e disponibilizando dados de forma que os programas que circulam na *web* sejam capazes de fazer inferência aos dados. É importante destacar que, para se ter a representação do conhecimento, precisa-se da Interoperabilidade Estrutural, Sintática e Semântica.

A Interoperabilidade Estrutural é aquela que permite que os dados sejam representados de forma diferenciada, contribuindo na especificação de tipos e possíveis valores para cada forma de representação. A Interoperabilidade Sintática está relacionada com regras precisas que permitem o intercâmbio de dados na *Web*. A Interoperabilidade Semântica, por fim, possibilita a compreensão e associação entre os dados.

Para tanto, são utilizadas as linguagens XML e RDF, que permitem expressar os dados para definir regras de raciocínio. O quadro 5 demonstra as características básicas de XML e TDF:

Quadro 5 - Características gerais da XML e RDF

XML (EXTENSIBLE MARKUP LANGUAGE)	RDF (RESOURCE DESCRIPTION FRAMEWORK)
Uma linguagem de representação de dados	É um modelo de dados para objetos (recursos) e relações entre eles.
Tem como foco a semântica dos dados representados e não sua forma de apresentação.	Uma linguagem usada para representar informações na Internet.
É uma linguagem de marcação extensível (<i>extensible Markup Language</i>) derivada do SGML.	São arquivos de dados ou metadados, tendo como um dos principais objetivos a criação de um modelo simples para armazenamento de informações.
Em XML, as <i>tags</i> não são pré-definidas.	É a base para a publicação e linguagem de dados.

Fonte: Elaborado pelos autores (2001)

2.3.2 A Camada Ontologia (*Ontology Layer*)

A camada ontologia (*ontology layer*) refere-se ao momento em que duas bases de dados utilizam terminologias diferentes para indicar a mesma informação, o que resulta na divergência de um mesmo conjunto semântico de dados. Também diz respeito ao momento em que uma mesma terminologia pode estar sendo usada com significados diferentes por aplicações distintas. Para resolver esses conflitos, usa-se a camada de ontologia que é capaz de definir qual mecanismo é estabelecido em um padrão das páginas de *web*.

2.3.3 A Camada Lógica (*Logic Layer*)

A camada *Logic Layer* tem como característica principal os possíveis relacionamentos de informação e as inferências de

conhecimento da *Web Semântica*. Faz-se importante destacar que as regras de inferência favorecem aos programas o sentido de raciocinar sobre os termos e seus significados.

No que se refere às implementações da *web semântica*, existem vários tipos de serviços que estão disponíveis nesse sistema, mas duas abordagens são importantes de se destacar: *Bottom Up* e *Top Down*. A primeira abordagem refere-se ao momento em que são inseridas *Tags* nos textos e dados para que as informações disponíveis sejam adaptadas a esse novo conceito. Na abordagem *Top Down*, torna-se mais importante o processamento dos dados existentes, assim como a criação de conexões e o estabelecimento do significado entre eles.

Existem outros serviços que estão em desenvolvimento ou em fase de testes, mas o importante é relatar que, de fato, a *Web Semântica* não possui uma forma definitiva de utilização; seu propósito é tornar as redes e resultados mais inteligentes.

A *Web Semântica* tem contribuído de forma significativa para a renovação dos tesouros, servindo como suportes para buscas semânticas e outros serviços. Hoje, as ferramentas que gerenciam os novos tesouros apresentam estratégias para que eles sejam criados, editados e consultados.

3 PROCEDIMENTOS METODOLÓGICOS

O corpus de amostra representativa desta pesquisa foi extraído da Base Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI), correspondendo ao período de 2015 a 08 de setembro de 2021, sobre o tema tesouro. Escolhemos os seguintes descritores: Tesouro, Vocabulário Controlado, Vocabulário de Palavras e Linguagem

Documentária. O levantamento da busca na base de dados da BRAPCI foi feito pelos seguintes campos: título, palavras-chave e resumo.

Conforme o levantamento dos descritores, tivemos uma amostra de 178 publicações sobre tesouros, das quais foram utilizadas 80, referentes apenas às tipologias que correspondem a artigos e artigos científicos, como estão caracterizados pela própria BRAPCI. Isto é, as outras 98 publicações, classificadas como resenhas, artigos de revisão, artigos incompletos, comunicação oral e outras denominações, foram descartadas, como se pode observar no Quadro 6. Incluem-se, ainda, nesse grupo, artigos duplicados e os que não correspondiam ao ano pesquisado, mas que estavam indexados na BRAPCI.

Quadro 6 - Quantitativo das Publicações pesquisadas

DESCRIPTORES UTILIZADOS	PUBLICAÇÕES (2015-2021)	PUBLICAÇÕES SELECIONADAS	PUBLICAÇÕES DESCARTADAS
Tesouro	76	33	43
Vocabulário Controlado	58	26	32
Vocabulário de Palavras	03	01	02
Linguagem Documentária	41	20	21
Totais	178	80	98

Fonte: BRAPCI (2021).

No que se refere ao caráter metodológico, este estudo é uma pesquisa caracterizada como descritiva, bibliográfica e documental. Escolhemos como percurso metodológico a bibliometria, sustentada na análise de produtividade de autores, o que converge com o objetivo principal da Lei de Lotka, que é levantar o impacto da produção de um autor numa área de conhecimento e, por isso, a escolha da bibliometria (OLIVEIRA, 1983).

Para a análise dos dados, foram observados dois critérios de dados. O primeiro refere-se aos aspectos extrínsecos numa análise bibliométrica, em que se identificaram os seguintes indicadores: ano de publicação, autoria, título do documento e publicação. Dessa forma, a bibliometria mapeia estatisticamente a estrutura dos conhecimentos em um determinado campo produzido, faz uso de métodos quantitativos e permite análises qualitativas em relação ao estudo dos padrões de comportamento desse conhecimento (ARAÚJO, 2006). O segundo critério volta-se aos aspectos intrínsecos, numa análise de conteúdo.

Geralmente, análises bibliométricas e de conteúdo são realizadas a partir de levantamento estatístico e frequência. O método da análise de conteúdo contribui com os procedimentos que possibilitam uma análise com base em interpretações dos indicadores, bem como oportuniza a observação da relação entre a produção técnico-científica sobre tesouro e os princípios teóricos para a gestão dos tesouros em que os documentos têm características semelhantes e homogêneas.

4 ANÁLISE DE DADOS

Para a análise dos aspectos extrínsecos, foram levantados os itens: base de dados, ano de publicação, autor, palavras-chave e publicação de revista.

Ano de publicação - No período de 2015 a 08 de setembro de 2021, observa-se uma diminuição de 50% nas publicações dos anos de 2015 e 2016 sobre o tema. Porém, a partir de 2017, vê-se um aumento, com estabilidade no ano seguinte, e uma tendência crescente de publicações sobre tesouro em 2019, sendo este o ano em que houve mais publicações. Os anos 2020 e 2021 apresentam novamente uma

diminuição no número de publicações, atingindo 23% da produção (Quadro 7).

Quadro 7 - Ano de Publicação

Ano	Quantidade de Artigos
2015	12
2016	06
2017	11
2018	11
2019	17
2020	13
2021	10
Total	80

Fonte: (BRAPCI, 2021)

Autores – Houve a ocorrência de 63 autores. Desse total, apenas 07 aparecem como único autor da produção sobre o tema; os demais apresentam em coautoria. Os autores com mais produção são Walter Moreira, Gercina Lima, Brisa Sousa, Luciana Davanzo Rita Laipelt, Michelly Vogel, Flavio da Silava, Benildes Maculan e Cibele Santos (Quadro 8).

Quadro 8 - Autores que mais publicaram

	Autor	Quantidade de publicação
1	MOREIRA, Walter	5
2	LIMA, Gercina Angela Borém Oliveira	4
3	FUJITA, Mariângela Spotti Lopes;	4
4	SOUSA, Brisa Pozzi de;	4
5	DAVANZO, Luciana;	3
6	LAIPELT, Rita do Carmo	3
7	VOGEL, Michelly Jabala Mamede	3
8	SILVA, Flávio Pacheco da	3
9	MACULAN, Benildes Coura Moreira dos Santos	3
10	SANTOS, Cibele Araújo Camargo Marques dos	3

Fonte: (BRAPCI, 2021)

Publicação - Foram selecionados e examinados 25 tipos de publicações de revistas, sendo que 06 revistas tiveram o menor número de publicação, enquanto uma revista obteve 19 publicações sobre o tema tesouro (Quadro 9):

Quadro 9 - Revistas que publicaram

	REVISTA	QUANTIDADE
1	Ágora	2
2	BIBLOS - Revista do Instituto de Ciências Humanas e da Informação	2
3	Ciência da Informação	2
4	DataGramZero	2
5	e-Ciencias de la Información (Costa Rica),	1
6	Em Questão	5
7	Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação,	5
8	InCID: Revista de Ciência da Informação e Documentação	5
9	Informação & Informação	19
10	Informação em Pauta	2
11	Informação@Profissões	1
12	Logeion: filosofia da informação	1
13	Memória e Informação	1
14	Múltiplos Olhares em Ciência da Informação	2
15	Páginas A&B, Arquivos e Bibliotecas (Portugal)	2
16	Perspectivas em Ciência da Informação	8
17	Ponto de Acesso	1
18	Prisma.com (Portugal)	2
19	Revista Analisando em Ciência da Informação	3
20	Revista Bibliomar	2
21	Revista Brasileira de Biblioteconomia e Documentação,	2
22	Revista Digital de Biblioteconomia & Ciência da Informação	5
23	Revista Folha de Rosto	1
24	Revista Ibero-Americana de Ciência da Informação	2
25	Revista P2P e INOVAÇÃO	2
	Total	80

FONTE: (BRAPCI, 2021)

Palavras-chave - Buscou-se identificar os termos mais frequentes no campo palavras-chave dos registros recuperados, assim como determinar as áreas às quais o estudo do tesouro pode estar relacionado. No total, foram computadas 315 palavras-chave, das quais as mais citadas foram: Vocabulário Controlado – 24 vezes; Organização do

Conhecimento – 12 vezes; Tesouro – 15 vezes e Sistemas de Organização do Conhecimento – 08 vezes.

A análise dos aspectos intrínsecos ou de conteúdo foi desenvolvida em: **Análise** - Foram selecionados os trabalhos que mencionam, em seus conteúdos, algum tipo de relação com a base teórica apresentada na revisão de literatura. O critério foi definido por meio de leituras dos campos título, palavra-chave e resumo. Com base nesse parâmetro, foram identificados 22 documentos para a análise temática e selecionados para a leitura na íntegra. A análise dos temas associados a tesouros sugere a tendência de estudos sobre o assunto no campo da Ciência da Informação. Organização do Conhecimento e Recuperação da Informação aparecem como os temas mais frequentes. Entre os que tiveram baixa frequência, destacam-se *web* semântica e *software* livre. Ademais, observou-se, em muitos artigos, a preocupação em definir tesouro para fins de recuperação da informação.

4 CONSIDERAÇÕES FINAIS

Quanto ao objetivo proposto pela pesquisa, evidenciou-se, com base na revisão da literatura da área, o tesouro enquanto linguagem documentária, elaborado com interface na terminologia. Isso possibilita uma representação bem mais adequada e compatível com a linguagem dos usuários, permitindo, assim, uma recuperação da informação precisa e pertinente. A construção de um tesouro consiste em uma atividade interdisciplinar, ligada, principalmente, aos campos da Ciência da Computação e da Ciência da Informação pelo fato de que a Ciência da Computação é constituída de tecnologia para implementar o uso de tesouros empregados na *Web Semântica*.

Dessa forma, foi possível identificar que os estudos da Organização do Conhecimento, no contexto da Ciência da Informação, vêm ganhando dimensões proporcionais, devido à aceleração do uso de tecnologias da *web*. Pesquisadores do campo da (OC) trabalham na busca de novas teorias e metodologias que ajudem no referencial teórico dos Sistemas de Organização do Conhecimento.

Com relação à análise de conteúdo, nossa principal dificuldade foi em relação ao número diversificado de temáticas nos artigos, o que gerou uma limitação dos resultados; porém, há uma considerável presença da Organização do Conhecimento como embasamento teórico do percurso escolhido pela maioria dos autores, o que fundamenta a consistência da gestão do tesouro enquanto instrumento de organização do conhecimento.

REFERÊNCIAS

AMERICAN NATIONAL STANDARD/NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies**. Bethesda: NISO Press, 2005.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. *In*: CARRARA, K. (org.). Educação, Universidade e Pesquisa. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001.

BARROS, T. H. B; SOUSA, B. T. R. R. Organização do Conhecimento e Arquivologia: abordagens metodológicas. **Informação & Informação**, Londrina, v. 24, n. 2, p. 76-92, nov. 2019. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/38290>
Acesso em: 25 set. 2020.

BERNERS-LEE, T. **Semantic web road map**. 2001. Disponível em: <http://w3.org/DesignIssues/Semantic.html>. Acesso em: 25 set. 2020.

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? *In*: LARA, M. L.G.; SMIT, J. (org.). **Temas de pesquisa em Ciência da Informação no Brasil**. São Paulo: Escola de Comunicação e Artes/USP, 2010.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, p. 348-358, 2006. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/273>. Acesso em: 6 set. 2020.

CERVANTES, B. M. N. **A construção de tesouros com a integração de procedimentos terminográficos**. 2009. 198f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências. Marília, 2009.

CERVANTES, B. M. N. **Contribuição para a Terminologia do Processo de Inteligência Competitiva**: estudo teórico e metodológico. Marília, 2004. 183f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências. Marília, 2004.

CINTRA, A. M. M. Elementos de linguística para estudos de indexação. **Ciência da Informação**, Brasília, v. 12, n. 1, p. 5-22, 1983.

CINTRA, A. M. M. Estratégias de leitura em documentação. *In*: SMITH, J. W. **Análise documentária**: a análise da síntese. 2. ed. Brasília: IBICT, 1989.

CINTRA, A. M.; TÁLAMO, M. F. G. M.; LARA, M. L. G.; KOBASHI, N. Y. **Para entender as linguagens documentárias**. São Paulo: Polis/APB, 1994.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115> Acesso em: 25 set. 2020.

DAHLBERG, I. *Knowledge organization: its scope and possibilities*. **Knowledge Organization**, Frankfurt, v. 4, n. 20, p.211-222, 1999.

DODOBEI, V. L. D. **Tesouro**: Linguagem de representação da memória documentária, Rio de Janeiro: Inter ciência, 2002.

GUIMARÃES, J. A. C. Organização do Conhecimento: passado, presente e futuro sob a perspectiva da ISKO. **Informação & Informação**, Londrina, v. 22, n. 2, p. 84-98, maio/ago. 2017. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/31443>

Acesso em: 25 set. 2020.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge organization**, Frankfurt, v. 30, n. 2, p. 87-111, 2003.

HJØRLAND, B. What is Knowledge Organization (KO)? **Knowledge organization**, Frankfurt, v. 35, n. 2/3, p. 86-101, 2008.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization**, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.