

RESUMO

O artigo demonstra a importância da representação no contexto da recuperação da informação, descrevendo as formas de estruturação do conhecimento registrado. Enfatiza a tarefa de incorporar o conhecimento subjetivo e tecnológico à realidade dos recursos, delineando a representação para muito além do seu valor instrumental obtido nos dados empíricos. A produção, organização, recuperação e uso da informação correspondente aos fluxos da informação encontram na Web semântica uma forma de gerir as representações temáticas e descritivas no âmbito tecnológico da codificação. Conclui destacando que os desafios e tendências das formas de representação no contexto da recuperação da informação estão seguindo uma linha evolutiva que começa com a aquisição do conhecimento e termina com as bases de conhecimento dos sistemas especialistas, na qual considera-se a Organização do Conhecimento como a espinha dorsal dos atos de representar, recuperar e democratizar o conhecimento.

Palavras-chave: Representação do Conhecimento. Recuperação da Informação. Sistema de Organização do Conhecimento. Necessidade do usuário.

ABSTRACT

This article demonstrates the importance of representation in the context of information retrieval, describing the ways of structuring registered knowledge. Emphasizes the task of incorporating the subjective and technological knowledge to the reality of the resources, delineating the representation far beyond its instrumental value obtained in the empirical data. The production, organization, retrieval and use of information corresponding to the information flows are in the Semantic Web a way to manage the thematic and descriptive representations in the technological scope of the encoding. Concludes highlighting that the challenges and trends of the forms of representation in the context of information retrieval are following an evolutionary line that begins with the acquisition of knowledge and ends with the knowledge bases of expert systems in which it is considered the Knowledge Organization as the backbone of efforts to represent, recover and democratize knowledge.

Keywords: Knowledge Representation. Information Retrieval. Knowledge Organization System. User needs.

¹ Programa de Pós-graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais.

A alta disponibilidade de aparatos tecnológicos acessíveis no meio virtual demanda ações que visem uma melhor estruturação do conhecimento para fins de recuperação da informação desejada. A representação do conhecimento procura ao mesmo tempo, dar visibilidade e relevância às vertentes necessárias para uma boa recuperação da informação que perpassam por estruturas adequadas durante a sistematização do conhecimento.

As resoluções pragmáticas da estruturação do conhecimento registrado possuem conjecturas intrínsecas que abarcam um conjunto de fenômenos e possibilidades lógicas para o estabelecimento de critérios na aderência à realidade observada. Neste sentido, a credibilidade e abrangência dos mapeamentos sistemáticos de um campo do conhecimento consolidam as alternativas potenciais e flexíveis de representação.

Durante a sistematização dos conceitos, a natureza do significado está associada a uma representação mental e a uma noção de referência. Então, isso significa que ao aprender o que as palavras expressam, pode-se determinar a posição do processo cognitivo e representacional em um determinado contexto. O sucesso da comunicação depende das similaridades encontradas nos construtos mentais e do compartilhamento dessas representações que codificam o conhecimento registrado.

Nas investigações científicas, o interesse pela eficácia e materialidade no âmbito da tecnologia partiu da necessidade em distinguir o conhecimento científico tecnológico da atividade tecnológica, onde o produto do conhecimento tecnológico é a atividade tecnológica caracterizada por tentativas e erros ao considerar os níveis de organização dos artefatos tecnológicos (TAKAHASHI, 2009). Para fins de comprovação, o processo de síntese e o uso das terminologias adequadas para a representação tecnológica estão ligadas à complexidade e entendimento das teorias e suas possíveis causas na determinação da natureza do conhecimento.

Nesta perspectiva, o objetivo desta pesquisa é demonstrar a importância da representação no contexto da recuperação da informação, possibilitando novas formas de estruturação do conhecimento registrado. A tarefa de incorporar o conhecimento subjetivo e tecnológico à realidade dos recursos, delinea a representação para muito além do seu valor instrumental obtido nos dados empíricos. Normalmente, as evidências empíricas destacam problemas de consistência teórica que necessitam de esforços complementares para identificar e formalizar as estruturas de conhecimento.

2 O CONTEXTO DA REPRESENTAÇÃO NA RECUPERAÇÃO DA INFORMAÇÃO

No contexto da representação, considera-se que os seres humanos são conhecedores e observadores da realidade externa, assimilando coisas (processo de conhecimento) por meio de suas representações na mente/cérebro do sujeito cognoscente que serão armazenadas e processadas por máquinas/computadores (ALVARENGA, 2003). O conhecimento relaciona-se com o objeto, enquanto representação e com o sujeito, sob a forma da consciência. Tem-se a consciência como uma condição universal de um conhecimento em geral, na qual a representação não é o conhecimento, mas o conhecimento pressupõe sempre a representação (GONZALEZ DE GOMEZ, 1993).

Nesta perspectiva, nascem os produtos da estruturação do conhecimento, os Sistemas de Organização do Conhecimento, que são representações do conhecimento registrado para posterior recuperação. Essas representações precisam passar por uma análise de domínio conceitual de acordo com sua realidade para que possa proporcionar maior visibilidade e harmonização destes conceitos. A análise de domínio é realizada a partir da observação da realidade estudada e atende a um objetivo específico da comunidade discursiva:

No caso da representação do conhecimento, a representação construída não se restringe ao conhecimento expresso por um autor, ela é fruto de um processo de análise de domínio e procura refletir uma visão consensual sobre a realidade que se pretende representar. A representação do conhecimento reflete um modelo de abstração do mundo real, construído para determinada finalidade (BRASCHER; CAFÉ, 2008, p.6).

A Representação do Conhecimento tem atividades nas áreas de Ciência da Informação, Ciência da Computação e na Terminologia. As contribuições de cada área para a RC são, respectivamente: linguagens documentárias, estruturas computáveis e sistematização de conceitos (CAMPOS, 2004). Neste contexto, entende-se que as três áreas do conhecimento estão interligadas por um mesmo objetivo: representar o conhecimento. A área de OC participa desta união para processar as estruturas de conhecimento e organizá-las de acordo com suas especificidades semânticas:

Representação do conhecimento (RC) é uma área de estudo estreitamente ligada à OC, em especial, no que tange à rede semântica. Esta representa o conhecimento na forma de nós (conceitos, atributos...) ligados por arcos (relacionamentos). Ao contrário dos tesouros, que empregam apenas termos (substantivos) e indicam somente a existência de relações entre eles no plano conceitual e no plano da língua, as taxonomias e ontologias – que também usam técnicas de classificação e tesouros – definem esses relacionamentos e podem incorporar adjetivos, expressões sintáticas e variações morfológicas, de sorte que possam ser usados pelo *software* no processamento das buscas. Essas são características específicas das ontologias (GOMES, 2009, p. 74).

No campo da Representação do Conhecimento, Gomes (2009, p. 78) descreve que a conceituação é composta pela “identificação e análise de diversos atributos / facetas dos objetos / entidades / conceitos que constituem o domínio em relação ao interesse dos usuários e que organizam atributos / facetas no interior do domínio em alguma sequência útil” e são abordadas na teoria do conceito de Dahlberg (1978) e nas relações semânticas. A autora aponta que há uma deficiência de estudos epistemológicos, principalmente nas áreas de conceituação de conceitos, objetos e relações e na classificação.

A estruturação e a representação do conhecimento facilitam a recuperação da informação adequada para uma necessidade de informação. Os fundamentos teóricos e metodológicos da Organização do Conhecimento (OC) dão subsídios para a construção dos Sistemas de Organização do Conhecimento (SOC) que por sua vez, asseguram os aspectos semânticos de uma área especializada e contribuem para a definição do seu campo nocional com relações e associações. Esses fundamentos são uma porta de entrada para um tratamento da informação mais eficaz e conseqüentemente, uma busca e recuperação mais satisfatória. Por isso, torna-se importante destacar as abordagens fundamentais de um Sistema de Recuperação da Informação:

Nos processos de tratamento e recuperação da informação, o conceito, tanto na etapa de análise do documento, quanto na tradução com fins de padronização, ocupa lugar central. Um sistema de recuperação de informação prevê duas abordagens fundamentais: a análise chamada “física” ou “descritiva” do documento (quando são descritos dados bibliográficos, como autoria, editora, número de páginas, ano de publicação, etc.); e a análise de “conteúdo” ou “temática” do documento (em que o conteúdo do documento é representado por meio de conceitos (termos) (KOBASHI; FRANCELIN, 2011, p. 11).

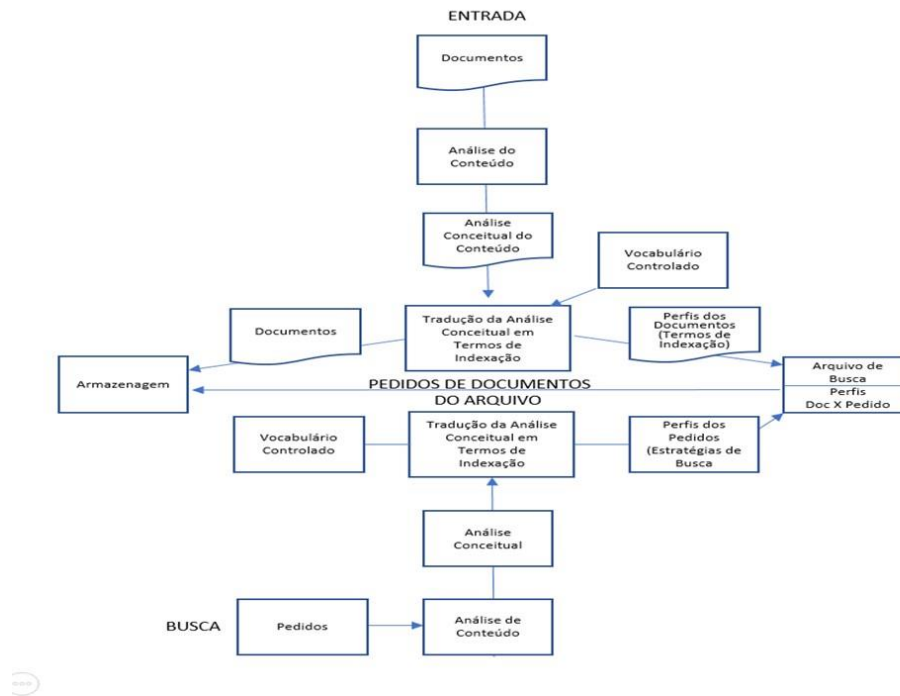
A Recuperação da Informação (RI) tem origem na Ciência da Computação e adentrou a Ciência da Informação justamente nos processos de busca e relevância dos documentos. Devido ao intenso número de documentação científica e técnica gerada durante o século XX, Calvin Mooers, em 1951, fundou o termo “Recuperação da Informação”, elegeu o termo “relevância” e o conceito de “descriptor” para o vocabulário controlado no processo de Recuperação da Informação (MOOERS, 1951). A RI teve início com “esquemas codificados para estruturar e representar conteúdos, como por exemplo, as fichas perfuradas na margem” (GOMES, 2017, p. 50).

Para Saracevic (1991, p.2) a Recuperação da Informação é conceituada “[...] como aquilo que abrange os aspectos intelectuais da descrição da informação e suas especificações para busca, e ainda quaisquer sistemas, técnicas ou máquinas que são utilizados para realizar a operação”. A comunicação entre um Sistema de Recuperação da Informação (SRI) e a necessidade de informação de um usuário surge a partir do momento em que se observa uma lacuna entre o conhecimento que o usuário obtém sobre o problema ou tópico e o que o usuário precisa saber para resolver o problema. Todo conhecimento acumulado, organizado e registrado é composto por uma junção de conceitos. Logo, esses conceitos precisam ser definidos por unidades terminológicas para que possam ser transferidos ao público interessado de forma adequada.

Quando um conceito é retirado de seu ambiente de uso social ou contextual e colocado no ambiente do sistema, ele passa de uma categoria discursiva, dinâmica e plurissignificativa, para outra, fixa e monossignificativa. Há, nesse movimento, perda de conteúdo semântico na transferência do contexto de uso do conceito para o contexto de controle do conceito para recuperação de informação (KOBASHI; FRANCELIN, 2011, p. 13).

O controle e padronização dos conceitos devem contribuir para a ocorrência de compatibilidade entre a linguagem dos especialistas, a linguagem dos SRI e a linguagem dos usuários para que haja o acesso a informação desejada. Nesse contexto, Cesarino e Pinto (1980, p.33) demonstram a importância do conceito no processo de entrada e saída do documento em um sistema de recuperação da informação (Ver Figura 3).

Figura 1 - Processo de entrada e saída do documento em SRI



Fonte: Cesarino e Pinto (1980)

A área de Recuperação da Informação (RI) cresceu consideravelmente nos últimos anos com a inserção da modelagem, classificação de textos, arquitetura de sistemas, *interfaces* de usuário, visualização de dados, filtragem e linguagem, os quais ultrapassaram os limites da indexação e da busca por documentos e pode ser estudada sob duas vertentes: uma centrada no computador e a outra centrada no usuário. Na visão centrada no computador, a RI é composta por construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento a fim de melhorar os resultados. Já na visão centrada no usuário, a RI estuda o comportamento do usuário e procura entender suas principais necessidades (BAEZA-YATES E RIBEIRO-NETO, 2013).

Observa-se que a desterritorialização do documento, a desvinculação da forma física e a ruptura da linearidade do acesso colaboraram para as mudanças no processo de recuperação da informação, exigindo novos métodos de busca e representação. De acordo com Chu (2010), os principais componentes de um Sistema de Recuperação da Informação (SRI) são: a base de dados (contém informação representada e organizada), a linguagem, o mecanismo de busca e a interface.

Baracho (2016) relata que os avanços da recuperação da informação caminham em duas direções. A primeira focaliza as formas de representação do conhecimento e a forma de enquadrar uma determinada busca dentro dos parâmetros pré-estabelecidos. A segunda foca na linguagem natural como forma de recuperar a informação e os sistemas buscam soluções para os problemas que chegam de diversas formas sem padronização ou formatação pré-estabelecida. Nesse propósito, considera-se que a base para os sistemas de recuperação da informação são os termos de indexação ou palavras-chave para simplificar os problemas de busca e representação.

A RI apresenta três modelos clássicos, denominados: booleano, vetorial e probabilístico. No modelo booleano, os termos de indexação não tem peso algum associado; são simplesmente elementos de um conjunto. Nos modelos vetorial e probabilístico, os termos de indexação possuem pesos associados com o objetivo de melhorar o ordenamento dos documentos (BAEZA-YATES E RIBEIRO-NETO, 2013).

O ordenamento dos documentos lista os documentos mais relevantes de acordo com o grau de similaridade entre eles durante a busca. Para isso, o sistema utiliza algoritmos preditivos que trabalham com premissas básicas de relevância e dão origem ao ranqueamento, ou seja, os documentos que aparecem no topo dos resultados são considerados mais relevantes. Os algoritmos têm a função de alcançar o grau de relevância considerado pelo usuário durante a busca, visto que uma mesma busca pode ter resultados diferentes para cada usuário. O uso da linguagem natural também ajuda a simplificar uma busca com seus resultados de acordo com o perfil do usuário.

Desta forma, para que o ranqueamento obtenha resultados mais próximos da necessidade do usuário, é necessário registrar os termos de indexação de forma correta, precisa e coesa, considerando também as relações entre os termos que são bastante discutidos no processo de organização do conhecimento. Porém, o ranqueamento baseado em texto não é suficiente para alcançar as necessidades do usuário, então é preciso considerar os links entre páginas Web e a interoperabilidade entre sistemas para melhorar a recuperação da informação desejada.

Sabe-se que os especialistas em informação são eficientes na manipulação das bases de dados e familiarizados com as políticas de vocabulário e indexação e, em geral, formulam melhores estratégias de busca e fazem uso mais eficiente do sistema. Porém, normalmente, os usuários não têm tempo para aprender como usar um sistema de maneira eficaz ou eficiente. Então há uma forte tendência em priorizar cada vez mais o uso comum em que os usuários sejam capazes de intervir na busca sem a intermediação de um especialista da área. Por isso, dá-se ênfase a design, avaliação e uso de sistemas *on-line* como aspectos que devem ser continuamente considerados em um SRI (LANCASTER; FAYEN, 1973).

Nesse contexto, o campo da organização do conhecimento está amadurecendo ao estabelecer novas formas de melhorar o registro do conhecimento nos sistemas com a finalidade de proporcionar um uso mais eficiente das bases de dados e obter respostas mais relevantes.

Muitos pesquisadores vêm discutindo e analisando possibilidades de adaptação, desenvolvimento e construção de teorias e métodos que forneçam bases mais consistentes para a organização de conceitos em sistemas de recuperação da informação. Podemos identificar na literatura brasileira, o “incômodo” permanente diante de questões cognitivas, sociais e culturais quando o assunto passa pelo cânone normativo de conteúdos documentais (KOBASHI, FRANCELIN, 2011, p. 20).

Um dos componentes que mais influenciam um bom SRI é a sua *interface*. Os critérios de usabilidade adotados para uma *interface*, trazem benefícios como otimização do tempo; facilidade de uso e aprendizado do sistema buscas mais confiáveis e rápidas (PEREIRA, 2018). O uso da categorização permite que o usuário encontre informações relevantes e úteis de acordo com os conceitos usados na área do conhecimento em que está sendo pesquisada. As experiências e o conhecimento adquirido acerca do assunto ou da área, também são premissas para uma boa recuperação da informação:

O processo de estruturação por meio da categorização facilita a recuperação da informação pelos usuários. Zins (2000) menciona que a recuperação da informação é um processo destinado a atender as necessidades de informação do pesquisador. Portanto, sua avaliação é principalmente com base em seu resultado. Uma análise fenomenológica da recuperação da informação revela pelo menos cinco noções básicas que afetam os resultados da pesquisa: o pesquisador e suas capacidades, expertise e experiência em pesquisa; o processo cognitivo e o comportamento de busca; o domínio da informação; a tecnologia da informação; e as informações necessárias (DIAS; ALMEIDA, 2017, p. 874).

Logo, a Recuperação da Informação mudou consideravelmente com a expansão da *Web* e com o advento de interfaces gráficas e dispositivos de armazenamento. Os usuários reais e potenciais estão cada vez mais ansiosos por informações relevantes e em tempo hábil para as necessidades emergenciais. Por isso, a organização do conhecimento disponibiliza os recursos necessários para tornar a recuperação da informação mais confortável para esses usuários, como por exemplo, os Sistemas de Organização do Conhecimento (SOC).

3 OS SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO NA CIÊNCIA DA INFORMAÇÃO

Atualmente, os estudos da área de Organização do Conhecimento (OC) na Ciência da Informação estão cada vez mais procurando novas formas de estruturar o conhecimento e uma delas é a implantação de Sistemas de Organização do Conhecimento (SOC) voltados, em linhas gerais, para a atualização das tradicionais tabelas de classificação, o desenvolvimento de novas metodologias e o conseqüente emprego em meio eletrônico (GOMES, 2009). Neste aspecto, a aproximação da Biblioteconomia e da Ciência da Informação com a Ciência da Computação está sendo cada vez mais necessária no âmbito da OC devido a inclusão das tecnologias.

As classificações, consideradas como um dos tradicionais Sistemas de Organização do Conhecimento (SOC), envolvem o julgamento do assunto, o que não pode ser totalmente efetivado por uma máquina, visto que apresentam os resultados do detalhamento e da cognição do conhecimento registrado. Os serviços e produtos da sociedade da informação e do conhecimento encontram contribuições teóricas e práticas no uso da classificação para explorar as tecnologias com mais afinco e criatividade. Gomes (2017, p. 61) complementa que, “não podemos nos esquecer que a Classificação é a espinha dorsal para a solução das questões semânticas, com larga tradição na Ciência da Informação”. Neste caso, o profissional da informação tem uma grande contribuição na complementação de registros automáticos. É importante mencionar que os esquemas influenciados por princípios filosóficos e os esquemas facetados estão cada vez mais sendo discutidos e trabalhados no campo da organização do conhecimento com o propósito da ordenação como atividade central.

As teorias e instrumentos da OC contribuem para a criação de ferramentas aplicáveis e estruturadas sistematicamente como resultado da interpretação de conteúdo. O campo da organização do conhecimento vem apresentando propostas de melhoria e métodos de padronização das informações registradas de forma sistemática e organizada, denominados de Sistema de Organização do Conhecimento (SOC). A Ciência da Informação aborda os SOC ou esquemas de representação do conhecimento como:

Instrumentos que fazem a tradução dos conteúdos dos documentos originais e completos, para um esquema estruturado sistematicamente, que representa esse conteúdo, com a finalidade principal de organizar a informação e o conhecimento e, conseqüentemente, facilitar a recuperação das informações contidas nos documentos (CARLAN; MEDEIROS, 2011, p. 55).

Complementando esta ideia, Hodge (2000, p. 78) classifica os Sistemas de Organização do conhecimento (SOC) como:

Todos os tipos de instrumentos usados para organizar a informação e promover o gerenciamento do conhecimento. Incluem os esquemas de classificação que organizam materiais em nível geral (como livros em estantes), cabeçalhos de assunto que provêm acesso mais detalhado e listas de autoridade que controlam versões variantes de chaves de acesso à informação (nomes geográficos e nomes de pessoas). Incluem, ainda, esquemas menos tradicionais, tais como redes semânticas e ontologias.

Os SOC são compostos por termos, definições, relacionamentos e propriedades dos conceitos com a finalidade de padronização terminológica. Essa padronização envolve 5 verbos-chaves: encontrar, identificar, selecionar, obter e navegar (SVENONIUS, 2000). A estruturação de um SOC pode ser simples ou multifacetada de acordo com as necessidades do campo do conhecimento. Os tradicionais sistemas de organização do conhecimento (tesauro, taxonomia, vocabulário controlado e sistemas de classificação) receberam um suporte tecnológico por meio do *World Wide Web Consortium* (W3C) denominado de *Simple Knowledge Organization System* (SKOS). Os SKOS possuem estruturas conceituais que utilizam as linguagens semânticas XML/RDF para publicar os termos definidos em vocabulários estruturados e formalizados e são usados para o tratamento e a recuperação da informação em buscas, mapeamentos e conexões entre os tipos de SOC.

Os SOC são usados pelas pessoas para encontrar informações e dar sentido e significado a elas por meio de programas de computador com estruturas significativas de conceitos formalizados. Eles são conhecidos como ontologia, tesauros, esquema de metadados, taxonomia, classificação, estrutura de diretórios da Web, plano de arquivamento, dicionário de sinônimos, glossário, mapa conceitual, folksonomia e muito mais. Muitos desses termos são usados em diferentes contextos e por diferentes comunidades para significar coisas diferentes e como os significados se sobrepõem (SOERGEL, 2014).

Para Carlan e Medeiros (2011, p. 54) os SOC possuem as funções de “eliminação da ambigüidade, controle de sinônimos ou equivalentes e estabelecimento de relacionamentos semânticos entre conceitos”. Porém, antes de desenvolver um SOC é necessário analisar as necessidades informacionais dos usuários do sistema e do campo de conhecimento, identificar qual será o tipo de SOC mais adequado para os usuários e em seguida, desenvolver um *hardware* e *software* coerentes com a arquitetura da informação e sua integração. As autoras destacaram alguns tipos de SOC que são frequentemente usados na Ciência da Informação no âmbito da organização da informação e do conhecimento: os sistemas de classificação, os tesauros, as taxonomias e as ontologias, conforme são descritas a seguir:

- **Sistemas de classificação bibliográfica:** São usadas normalmente em bibliotecas ou bases de dados bibliográficas. A Classificação Decimal de Dewey (CDD), a Classificação da Biblioteca do Congresso dos EUA (LC), a Classificação Decimal Universal (CDU), a

Classificação de dois pontos e a Classificação de Bliss são exemplos de classificações bibliográficas existentes no cenário da Biblioteconomia e da Ciência da Informação. De modo geral, as atividades rotineiras de um indivíduo requerem constantemente um processo de classificação, seja em seus objetos, seja em sua mente. Todos usam a separação por categorias e níveis de organização de acordo com a diversidade de seus elementos com o propósito de recuperar e representar o conhecimento.

- **Tesauros:** São instrumentos de padronização terminológica e consolidados por normas internacionais com termos específicos de uma área de conhecimento, usados por indexadores para representar um documento para fins de recuperação da informação desejada. São estruturados pelo conceito, termo, categorias e faceta elaborados durante as atividades práticas de planejamento, coleta de termo, controle terminológico, estabelecimento de relações entre conceitos e formas de divulgação e publicação.
- **Taxonomias:** As taxonomias fazem uso, primeiramente, da classificação entre características de um objeto e conceito em um domínio do conhecimento e são posteriormente divididas em categorias e subcategorias com vários níveis de integração. Possibilitam a criação de metadados e termos comuns de um domínio com o objetivo de recuperar as informações de forma mais rápida e precisa. São usadas em mineração de dados na Web e possuem flexibilidade às mudanças de conteúdo apresentadas pela realidade observada e assim atender as necessidades dos usuários.
- **Ontologias:** As ontologias são muito utilizadas também em outras áreas como a Ciência da Computação e a Medicina por promoverem a interoperabilidade entre os sistemas de informação por meio do compartilhamento e reutilização do conhecimento no contexto da Web semântica. São formadas por termos, definições e relações, elementos estes que são comuns em outros tipos de SOC e que possibilitam as inferências com o uso das linguagens OWL (*Web Ontology Language*), XML (*eXtensible Markup Language*) e XML Scheme, RDF (*Resource Description Framework*) e RDF Scheme. As ontologias têm se destacado na organização do conhecimento pelo fato de sistematizarem o conhecimento já estruturado para a obtenção de resultados mais relevantes e úteis.

Gomes (2009, p. 74) enfatiza que há uma preocupação com “as questões semânticas, terminológicas, conceituais e de relacionamento entre conceitos nas atividades de Biblioteconomia e Ciência da Informação” que norteiam a construção de um Sistema de Organização do Conhecimento (SOC). A autora cita ainda que o método de faceta é o mais citado na literatura para a construção de SOCs por produzir “classificação coerente, intuitiva, auto explicativa, além de permitir acesso múltiplo” e estabelecer conexões entre os fenômenos e as teorias epistemológicas. O interesse das pesquisas de OC em criar classificações com o método facetado tem influenciado a disponibilização de estruturas sistematizadas de conhecimento para fins de recuperação e acesso digital, tais como as taxonomias e as ontologias.

A ontologia classifica os seres e as coisas no âmbito da sua existência, observando suas características primárias e estruturando-os de forma hierárquica. A contextualização das relações resulta na formação de um modelo conceitual. A vantagem na utilização das ontologias é poder deslocar continuamente as hierarquias conforme as adaptações e mudanças conceituais. Além de representarem a realidade objetiva de uma área, ainda apresentam um grau de relacionamento elevado que proporciona uma maior representação do conhecimento em formato digital

composta por taxonomias e hierarquias em sua estrutura. Podem ser classificadas quanto à função, grau de formalismo, aplicação, estrutura e conteúdo (ALMEIDA; BAX, 2003).

O desenvolvimento e a classificação nas ontologias demandam definição de classes e conceitos, arranjo numa hierarquia (taxonomia), definição das propriedades e valores para que se possa corretamente preencher os valores no sistema. O planejamento de uma ontologia deixa, então, evidente, o processo de classificação (GOMES, 2009, p. 77).

Na Ciência da Informação, o termo ontologia é usado para descrever uma base de conhecimento. Tanto na Ciência da Computação quanto na Ciência da Informação, vemos a construção de conjuntos de termos (conceitos) que são usados em uma comunidade específica - ou domínio - mapeados em conjunto com os relacionamentos entre eles. O mapa resultante é chamado ontologia de um domínio. Nas ontologias da Ciência da Informação, estudam-se pistas para a evolução do conhecimento entre domínios. Em ontologias de Ciência da Computação, são usadas para gerar aplicativos centrados no domínio. E nas ontologias de gestão do conhecimento são usadas para impulsionar a inteligência competitiva, alimentando em um domínio o próprio conhecimento do grupo sobre si mesmo (SMIRAGLIA, 2014).

Almeida (2013) faz um esclarecimento necessário sobre as ontologias e seu uso na Filosofia, na Ciência da Computação e na Ciência da Informação, ilustrando essa distinção no quadro 3. Apresenta a ontologia na filosofia como uma disciplina filosófica que trata da natureza e organização da realidade enquanto que a ontologia na Ciência da Computação e na Ciência da Informação referem-se a um determinado objeto, um artefato que pode ser um sistema conceitual informal, um relato semântico formal, uma especificação de uma conceituação ou uma representação de um sistema conceitual.

Quadro 1 - Síntese sobre ontologias

Distinção	Campo	O que é?	Propósito	Exemplo
Ontologia como uma disciplina	Filosofia	Ontologia como um sistema de categorias	Entender a realidade, coisas que existem e suas características	Sistemas de Aristóteles, Kant e Husserl
Ontologia como um artefato	Ciência da Computação	Ontologia como uma teoria (lógica);	Entender um domínio e reduzir isto para um modelo;	BFO, DOLCE ² (genérico);
		Ontologia como um artefato (software)	Criar vocabulários para representação em sistemas e inferências gerais	OWL (linguagem KR) ³
	Ciência da Informação	Ontologia como uma teoria (informal);	Entender um domínio e classificar termos;	Sistema de Classificação de Ranganathan;
		Ontologia como um sistema conceitual informal	Criar vocabulários controlados para recuperação da informação de documentos	Catálogo, glossário, tesouro

Fonte: Almeida (2013, p. 1690)

² BFO - Basic Foundational Ontology; DOLCE - Descriptive Ontology for Linguistic and Cognitive Engineering.

³ KR - Knowledge Representation; OWL - Web Ontology Language

Almeida (2013) menciona que a ontologia não é um sistema de classificação ou um vocabulário controlado no sentido de que os vocabulários controlados são usados na Ciência da Informação como um conjunto normalizado de termos, mas uma ontologia é um tipo de vocabulário controlado quando se especifica o conjunto de predicados que podem ser usados para fazer representações sobre um domínio. O autor reforça que a Ciência da Computação usa ontologia para categorizar o mundo, mas enfatiza o processo de raciocínio. No entanto, a principal preocupação da Ciência da Informação é lidar com documentos que descrevem as entidades do mundo em diferentes domínios e de diferentes pontos de vista, não com o raciocínio lógico. Explicita ainda que a ontologia em Ciência da Computação é usada para se referir tanto a um vocabulário em uma linguagem de representação do conhecimento quanto a um tipo de teoria onde se explicam fenômenos, usando fatos e regras. O primeiro uso corresponde ao software que deve ser desenvolvido de acordo com diretrizes de implementação. Já o segundo uso retém a noção de ontologia da filosofia, ou seja, de um inventário de coisas no mundo e das relações entre elas em um domínio particular, dando-lhe o poder da interdisciplinaridade dos campos.

Na Ciência da Informação, os princípios ontológicos (baseados na abordagem filosófica) podem ser usados para apoiar a construção de estruturas categóricas para a representação do conteúdo dos documentos e também para conectar diferentes campos de pesquisa em um nível interdisciplinar. Para isso, deve-se considerar primeiro, os princípios ontológicos a priori, dedicado a estabelecer que tipos de coisas possam existir e coexistir no mundo e depois estabelecer que tipo de coisa existe de acordo com evidências empíricas, o que ocorre comumente na CI. Os resultados empíricos é que vão determinar qual a melhor abordagem, pois esses sistemas podem ter uma inclinação realista, cognitiva ou descritiva e podem não compartilhar os mesmos princípios. A literatura sobre ontologias em Ciência da Informação pode ser construída em pesquisas interdisciplinares e com base na fundamentação de investigações empíricas (ALMEIDA, 2013).

O trabalho na Ciência da Informação não se limita à representação do conteúdo, mas abrange a representação como um todo, geralmente a partir da perspectiva de uma determinada comunidade de usuários. Os profissionais da informação (bibliotecários) colaboram na construção das ontologias durante a identificação de conceitos e relacionamentos, visto que possuem uma base teórica e prática da Teoria da Classificação. A ontologia, nesse sentido, é um tema frutífero na pesquisa em CI, porque possui o potencial de explicar não só questões relativas ao conteúdo de um documento, mas todo o ambiente social envolvido na análise realizada por cientistas da informação. Deveríamos ver ontologias (como artefatos) como repositórios de conhecimento científico que são um objeto de interesse para a CI e poderiam ser estudadas como parte de um fenômeno de comunicação. Desta forma, o desenvolvimento de ontologias formais (linguagem lógica) pode produzir resultados que podem ser posteriormente aplicados em vocabulários controlados que lidam com a linguagem natural (ALMEIDA, 2013).

Portanto, pode-se inferir que os Sistemas de Organização do Conhecimento (*Knowledge Organization Systems* - KOS) são estruturados com conceitos, categorias, classes, relacionamentos entre eles e termos ou outras designações para esses conceitos e relacionamentos. Fornecem uma estrutura ou esquema para armazenar e organizar dados, informações, conhecimento sobre o mundo e sobre pensamentos - sobre o que é ou é suposto ser, o que poderia ser ou deveria ser - para compreensão, recuperação ou descoberta, além de possuírem muitas finalidades (SOERGEL, 2014):

- Apoiar o pensamento, a tomada de decisão, a integração do conhecimento e a descoberta de novos conhecimentos e lacunas;
- Suporte ao design e tomada de decisão;
- Fornecer um roteiro semântico para um indivíduo ou promover o entendimento compartilhado;
- Melhorar a comunicação em geral;
- Apoiar a aprendizagem e assimilar a informação;
- Apoiar boas pesquisas e práticas;
- Fornecer classificação para ação e para fins sociais e políticos;
- Dar suporte à descoberta, recuperação e análise de informações e filtragem, roteamento, organização e acompanhamento de produtos e serviços;
- Dar suporte a exibição significativa e bem estruturada de informações (na recuperação e aprendizado de informações);
- Dar suporte ao processamento de informações após a recuperação;
- Fornecer a base conceitual para sistemas baseados em conhecimento. Inteligência Artificial (AI);
- Modelagem de dados;
- Dar suporte para a interoperabilidade sintática e semântica e preservação de significado ao longo do tempo;
- Fornecer informações sobre termos, conceitos e outras entidades para leitores, escritores e tradutores;
- Processamento de linguagem natural.

Para desenvolver um Sistema de Organização do Conhecimento (SOC) eficaz é preciso, primeiramente, compreender todas as suas funcionalidades, aplicações e tipologias para verificar qual será a sua utilidade de acordo com a comunidade de usuários e assim obter os benefícios da representação do conhecimento. O uso adequado dessas finalidades só será possível mediante uma boa construção de um instrumento de organização do conhecimento. Nesta concepção, torna-se essencial obter orientações primárias quanto aos procedimentos práticos e emergentes. A norma ISO 25964-1/2 publicada nos anos de 2011 e 2012 orienta a construção de tipologias de SOC:

É apresentada em duas partes: a parte 1, publicada em 2011, trata das recomendações relativas ao desenvolvimento e manutenção de tesouros monolíngues ou multilíngues, incluindo formatos e protocolos para intercâmbio de dados; a parte 2 trata da interoperabilidade entre diferentes tesouros e outros tipos de vocabulário estruturado, como sistemas de classificação, listas de nomes de autoridade, ontologias etc., que não eram cobertos anteriormente por nenhuma norma internacional (LARA, 2014, p. 101).

A primeira parte da norma apresenta o conceito que está representado pelo termo, o qual constitui a estruturação de um vocabulário com as possíveis conexões e equivalências entre eles. Essas relações são descritas conforme a especificidade do relacionamento entre os termos. A norma aborda ainda a análise facetada com fins hierárquicos de estruturação dos termos. A segunda parte, apresenta a importância do mapeamento dos SOC e a interoperabilidade entre eles. Descreve os elementos principais dos vocabulários estruturados com exemplos de termos e definições.

O principal objetivo dessa norma é promover a interoperabilidade entre os SOC, considerando o mapeamento, a precisão e a formalização como aspectos primordiais para a ampliação dos termos. Quanto maior a interoperabilidade entre os sistemas, maiores as chances de recuperar a informação desejada, pois as conexões e equivalência entre os termos possuem um alto índice de representação. A ação chave da interoperabilidade é o mapeamento dos vocabulários. A interseção entre os conceitos durante o mapeamento, possibilitam a ampliação ou permuta dos termos que representam os conceitos, permitindo uma aprimoração e extensão dos vocabulários.

A norma ISO 25964-2/2012 dispõe de 3 modelos para a realização de mapeamentos:

Modelo de unidade estrutural (*structuralunity*) quando os vocabulários compartilham exatamente a mesma estrutura hierárquica e associativa (por exemplo, para o mesmo instrumento em diferentes línguas); modelo de ligação direta (*direct-linked*) para estabelecer ligações entre dois ou mais vocabulários que não compartilham a mesma estrutura; estrutura com centro (*hub structure*) para estabelecer as correspondências de cada vocabulário em relação a um vocabulário de referência (LARA, 2014, p. 101).

Cada tipologia de Sistema de Organização do Conhecimento (SOC) possui um modelo de representação por meio dos conceitos. Todos têm como foco o mapeamento das hierarquias, equivalências e associações possíveis entre os conceitos. A norma impulsiona o uso da codificação para facilitar a interoperabilidade, apoiando-se na mútua exclusividade e na hierarquização das estruturas. Com isso, o mapeamento dá a oportunidade de verificar os resultados e as comparações que irão ajudar na elaboração de novos instrumentos de organização.

É curioso perceber que as tipologias de SOC têm como resultado a apresentação do estado da arte sobre um determinado tema ou campo do conhecimento. Além das iniciativas pragmáticas da organização do conhecimento, o que está em construção atualmente, e ainda carece de pesquisas, são as investigações epistemológicas em conjunto com a representação do estado da arte proporcionado pelo SOC. É necessário que alcancem a visibilidade e ultrapassem as fronteiras, permitindo a representação da evolução do tema. Desta forma, as ideias plurais são bem-vindas na interação da epistemologia com a pragmática da OC por meio de um processo rizomático do conhecimento. A epistemologia agrega a historicidade e as evoluções sociais, culturais e técnicas de uma área com seus nós, complexidades e ligações.

Observa-se que a base da construção de tipologias de SOC considera a ampliação dos conceitos adotados no mapeamento, visto que as estruturas semânticas permitem a ocorrência de significados comuns entre significantes diferentes. A abertura dos conceitos está relacionada com seus contextos e significados transversais que possam seguir também uma estrutura hierárquica sem comprometer o significado do mapeamento. Assim, visualiza-se a oportunidade de criar um modelo conceitual sem exclusão de conceitos e sim agregá-los. É interessante construir instrumentos que demonstrem a evolução dos conceitos e suas interações teóricas e comunicativas, tendo em vista as mudanças de postura epistemológica, social e cultural no âmbito da produção e complexidade do conhecimento registrado.

Contemplar a complexidade, mais evidente quando enfrentamos problemas de organização de fenômenos das ciências sociais e das humanidades, exigiria novas lentes, bem como novas ferramentas. Os mapeamentos semânticos podem ser uma

dessas ferramentas já que, apesar de ainda serem construídos tendo como referência princípios clássicos de organização, permitem trazer à tona uma série de KOS cujo conhecimento construído pode ser consultado, escolhido, combinado e recombinao sem se prender aos vértices que estão em sua origem, apesar das lacunas relativamente às soluções locais que não estão visíveis por problemas de diversas ordens (LARA, 2014, p. 105).

As novas ferramentas de organização do conhecimento têm como base, os princípios elementares dos tradicionais procedimentos de organização e devem acompanhar as reflexões oriundas da produção do conhecimento para a construção de novos códigos de representação. Esses códigos apresentam categorias que permitem uma boa recuperação da informação. Entende-se que uma informação é considerada relevante para um usuário quando apresenta as seguintes características: precisão, completude, economia, flexibilidade, confiabilidade, simplicidade, atualidade e verificabilidade (STAIR, 1998).

4 CONSIDERAÇÕES FINAIS

As reflexões deste artigo demonstraram que a falta de coerência entre a organização do conhecimento e a sua representação no sistema pode dificultar a recuperação da informação, por isso é necessário ter o controle das informações que são armazenadas no sistema de RI. A representação precisa ser sintética e sistemática para facilitar a busca pelo usuário da informação. Extrair os descritores corretos que representam o conteúdo de um documento é uma ação chave para o sucesso da recuperação da informação.

A sistematização cognitiva do conhecimento e a organização dos conceitos fazem parte do processo de construção de Sistemas de Organização do Conhecimento (SOCs). Esses sistemas, uma vez concluídos e prontos para serem operados, permitem a interoperabilidade entre os sistemas de recuperação da informação. Dessa forma, facilita a busca e interação do usuário no momento da localização da informação desejada.

A organização do conhecimento está preocupada com a qualidade dos instrumentos desenvolvidos e com a sua representação no meio social e na comunidade científica. A organização cognitiva e social do conhecimento é fundamental na contextualização da realidade estudada por meio da exploração do material e do tratamento dos resultados obtidos por meio de inferências e interpretações adequadas. Os procedimentos sistemáticos e objetivos da representação dão suporte para uma recuperação eficiente e eficaz em um SRI. Ao longo do texto, vimos que a classificação tem um papel central nas atividades da organização do conhecimento que tem participação também nos processos de gestão do conhecimento, na arquitetura da informação, na criação de instrumentos semânticos e metadados necessários para uma boa disponibilização das informações ao usuário.

A produção, organização, recuperação e uso da informação correspondente aos fluxos da informação encontram na Web semântica uma forma de gerir as representações temáticas e descritivas no âmbito tecnológico da codificação. A utilização da Web semântica melhora a qualidade e interoperabilidade da recuperação da informação, possibilitando padrões, estruturas e linguagens universais que atendam as necessidades do usuário. A ontologia, por exemplo, se caracteriza atualmente como uma área de pesquisa emergente na Ciência da Informação e na sociedade contemporânea, pois permite que as comunidades de usuários compartilhem

informações por meio de sistemas automatizados e ajuda no entendimento de sua conceitualização e relações dentro da realidade observada de um domínio específico do conhecimento. Por estes motivos, entende-se que compartilhar o conhecimento e simplificar os conteúdos tornam a representação indispensável no processo de recuperação da informação.

Os desafios e tendências das formas de representação no contexto da recuperação da informação estão seguindo uma linha evolutiva que começa com a aquisição do conhecimento e termina com as bases de conhecimento dos sistemas especialistas, na qual considera-se a Organização do Conhecimento como a espinha dorsal dos atos de representar, recuperar e democratizar o conhecimento.

REFERÊNCIAS

ALMEIDA, Mauricio Barcellos. Revisiting Ontologies: a necessary clarification. **Journal of the American Society for Information Science and Technology**, v. 64, n. 8, p. 1682-1693, 2013.

ALMEIDA, Maurício Barcellos; BAX, Marcello Peixoto. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, DF, v. 32, n. 3, p. 7 – 20, set./dez. 2003.

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 8, n. 15, p. 18-40, 2003.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação: conceitos e tecnologia das máquinas de busca**. Porto Alegre: Bookman Editora, 2013.

BARACHO, Renata Abrantes. Organização e recuperação da informação pilares da arquitetura da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 9, n. 1, 2016.

BRÄSCHER, Marisa; CAFÉ, Lígia. Organização da informação ou organização do conhecimento. In: IX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: [s.n], 2008. p. 1-14. Disponível em: <[http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89\(2008\)-1835.pdf](http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89(2008)-1835.pdf)>. Acesso em: 05 abr. 2016.

CAMPOS, M. L. de A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ci. Inf.**, Brasília, v.33, p.1, p.22-32, jan./abr. 2004.

CARLAN, Eliana; MEDEIROS, Marisa Bräscher Basílio. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, [S.l.], v. 4, n. 1, 2011.

CESARINO, M. A. N. B.; PINTO, M. C. M. F. Análise de assunto. **Revista de Biblioteconomia de Brasília**, v. 8, n. 1, p. 32-43, 1980. Disponível em: <http://www.brapci.inf.br/index.php/article/download/16905>. Acesso em: 22 Jul 2018.

CHU, H. **Information Representation and Retrieval in the Digital Age**. 2. Ed. Medford, N.J.: Information Today, 2010.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, DF, v. 7, n. 2, dez. 1978. Disponível em: < <http://revista.ibict.br/ciinf/article/view/115/115>>. Acesso em: 17 jun. 2017.

DIAS, Guilherme Ataíde; ALMEIDA, Josiana Florêncio Vieira Régis de. A categorização como organização do conhecimento na perspectiva de ChaimZins. In: **Tendências atuais e perspectivas futuras em organização do conhecimento**. Centro de Estudos Interdisciplinares do Século XX, 2017. p. 869-878.

GOMES, Hagar Espanha. Tendências da pesquisa em organização do conhecimento. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1, 2009.

GOMES, Hagar Espanha. Marcos históricos e teóricos da organização do conhecimento. **Informação & Informação**, [S.l.], v. 22, n. 2, p. 33-66, out. 2017. ISSN 1981-8920. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/31442>. Acesso em: 27 maio 2019.

GONZÁLEZ DE GOMEZ, Maria Nélide. A representação do conhecimento e o conhecimento da representação: algumas questões epistemológicas. **Ciência da Informação**, v. 22, n. 3, p.217-222, 1993. Disponível em: <http://revista.ibict.br/ciinf/article/view/479>. Acesso em: 18 jul. 2017.

HODGE, Gail. **Systems of knowledge organization for digital libraries: beyond traditional authority files**. Washington, DC: Digital Library Federation, Council on Library and Information Resources, 2000.

KOBASHI, Nair Yumiko; FRANCELIN, Marivalde Moacir. Conceitos, Categorias e Organização do Conhecimento; Conceptos, Categorías y Organización Del Conocimiento. **Informação & Informação**, v. 16, n. 2, p. 1-24, 2011.

LANCASTER, F. W.; FAYEN, E. G. **Information Retrieval On-line**. Los Angeles, CA: Melville Publishing Co., 1973. 417 p.

LARA, Marilda Lopes Ginez de. Documentary language and terminology. **Transinformação**, v. 16, n. 3, p. 231-240, 2004.

MOOERS, C. N. Zatocoding applied to mechanical Organization of Knowledge. **American Documentation**, v. 2, n. 1, p. 20-32, 1951.

PEREIRA, Fernanda. Usabilidade em bibliotecas digitais. IN: LIMA, Gercina Ângela de (Org.). **Bibliotecas digitais: novas tendências na navegação em contexto**. Rio de Janeiro: Interciência, 2018.

SARACEVIC, T. Information science: origin, evolution and relations. In: VAKKARI, P.; CRONIN, B. (Eds). *Conceptions of library and Information Science*. COLIS Conference Tampère, 1991, Finland. **Proceedings...** Los Angeles: Taylor Graham, 1992.

SMIRAGLIA, R. P. **The elements of knowledge organization**. Switzerland: Springer International Publishing, 2014.

SOERGEL, Dagobert. Knowledge Organization Systems. Overview. 2014. Disponível em: www.dsoergel.com/SoergelKOSOverview.pdf. Acesso em: 14 maio 2018.

STAIR, Ralph M. **Princípios de sistemas de informação: uma abordagem gerencial**. 2. ed. Rio de Janeiro: LTC, 1998.

SVENONIUS, Elaine. **The intellectual foundation of information organization**. Cambridge, MA: MIT Press, 2000.

TAKAHASHI, Ricardo H. C. **A estrutura do conhecimento tecnológico do tipo científico**. Belo Horizonte: Editora UFMG, 2009.

ZINS, Chaim. Success, a structured search strategy: Rationale, principles, and implications. **Journal of the American Society for Information Science**, v. 51, n. 13, p. 1232-1247, 2000.