

## 1 INTRODUÇÃO

Os processos de representação e recuperação de informações sempre estiveram diretamente relacionados com as tecnologias disponíveis. A partir do desenvolvimento tecnológico das últimas décadas, novas formas de representar e recuperar informações passaram a fazer parte do cotidiano de toda sociedade.

McLuhan (1974) afirma que as tecnologias ampliam e estendem nossos sentidos e capacidade de compreensão, possibilitando que as utilizemos como uma extensão de nosso próprio corpo. Ainda segundo o autor, a roda, por exemplo, pode ser considerada como uma extensão de nossos pés. Do mesmo modo, podemos considerar o celular como uma extensão potencializadora de nossos ouvidos – motivo pelo qual nos sentimos “amputados” na sua ausência.

Entre os inúmeros avanços impulsionados pela era digital, o ambiente *Web* apresenta-se como uma realidade que tem assumido um papel preponderante, influenciando o modo como convivemos em sociedade, trabalhamos, aprendemos, compramos, nos divertimos e relacionamos.

Apresenta-se como desafio para a Ciência da Informação desenvolver ferramentas, métodos e teorias que possibilitem uma melhor compreensão e utilização do ambiente *Web*. Contudo, devido à abrangência e à diversidade de questões envolvidas nesta discussão, torna-se complexa a tarefa de analisar as novas tendências de pesquisa envolvendo *Linked Data*, *Data Science*; Publicação Ampliada e *Web Semântica*.

O presente artigo considera o âmbito dos Repositórios Científicos digitais como um microcosmo de aplicação das atividades relacionadas à Representação e à Recuperação, desenvolvidas no campo da Ciência da Informação, no contexto *Web*, e expõe uma síntese do trabalho apresentado no Seminário *Representação e Recuperação da Informação na Web: aspectos teóricos e tecnológicos*, realizado no âmbito do XIX Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), no Grupo de Trabalho “Estudos Históricos e Epistemológicos da Ciência da Informação (GT-1)”, no dia 24 de outubro de 2018.

Consoante com a disposição do GT-1 de promover uma maior articulação entre os Grupos de Trabalho vinculados à Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (Ancib), valorizando experiências de pesquisas integradas pela temática, e no intuito de compartilhar “estudos relacionados aos processos, produtos e instrumentos de representação do conhecimento (aqui incluindo o uso das tecnologias da informação) e as relações inter e transdisciplinares neles verificadas”, membros permanentes do GT 02 “Organização e Representação do Conhecimento” apresentam estudos desenvolvidos no

<sup>1</sup> Seminário realizado no âmbito das atividades do GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação, durante o XIX Encontro Nacional de Pesquisa em Ciência da Informação, Londrina, 2018.

contexto da Representação e Recuperação da Informação por meio de termos, em ambientes especializados, neste estudo, os periódicos científicos eletrônicos.

## 2 REPRESENTAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

O estudo da Representação e Recuperação da Informação considera, em um de seus vieses investigativos, o aprimoramento de etapas da construção de instrumento de Representação do Conhecimento, aliado a contribuições de procedimentos para a aplicação de tecnologias da informação, também produz uma representação de conceitos, por meio de termos, tendo em vista a obtenção de um vocabulário consistente, que compõe a base para a representação e a recuperação temática da informação e é compatível com a demanda de áreas de especialidade.

A origem do termo representação decorre da forma latina *repraesentare* – isto é: “[...] fazer presente ou apresentar de novo. Fazer presente alguém ou alguma coisa ausente, inclusive uma ideia, por intermédio da presença de um objeto” (MAKOWIECKY, 2003, p. 3).

A definição de representação mais ampla foi dada por Palmer (1978, p. 262), que afirma: “Uma representação é, antes de mais nada e sobretudo, algo que significa algo mais. Em outras palavras, é uma espécie da coisa (ou das coisas que representa)”.

Novellino (1998, p. 137) define a representação da informação como:

A substituição de uma entidade lingüística longa e complexa - o texto de um documento - por sua descrição abreviada. Sua função é demonstrar a essência do documento. A representação da informação é um processo primeiro da transferência da informação e necessário para enfatizar o que é essencial no documento, considerando sua recuperação.

Marcondes (2001, p. 64), ao discorrer sobre a representação, infere que ela é “[...] um processo ocorrendo na mente de alguém, produzindo nesta mente algo distinto do objeto a que se refere”. O autor ainda salienta que a representação deverá “ser suficientemente rica sob o aspecto cognitivo e, ao mesmo tempo, sintética para economizar a energia do usuário de uma maneira significativa” (MARCONDES, 2001, p. 67).

As autoras Lima e Alvares (2012, p. 01) afirmam que “representar é o ato de utilizar elementos simbólicos – palavras, figuras, imagens, desenhos, mímicas, esquemas, entre outros – para substituir um objeto, uma ideia ou um fato”.

Nesse contexto, a representação é realizada visando à comunicação e à recuperação da informação, assim, é necessário que seja realizada de maneira compreensível, contextual e verossímil.

Nesse sentido, o processamento eficiente da informação disponível torna-se a cada dia um importante fator para que ela possa expressar o seu valor e, por conseguinte, ser recuperada, acessada, utilizada e apropriada por quem dela necessite.

Nesse momento é importante destacar que a maneira que será definida a representação deverá ser apropriada às tarefas que os usuários realizam (PESCHL, 2002), visto que são os usuários que se apossarão e usarão as representações para que seja possível a Recuperação da Informação.

O ano de 1945 é frequentemente referenciado na literatura como um marco inicial para o campo da Recuperação da Informação, a partir da divulgação do célebre artigo *As We May*

*Think*, publicado por Vannevar Bush. Neste artigo, Bush descreveu uma máquina que batizou como Memex, o trabalho costuma ser indicado como inspiração e ponto de partida dos desenvolvimentos subsequentes relacionados aos computadores digitais (BUSH, 1945; BUCKLAND, 1992).

Ao ler o artigo supracitado, e considerando o avanço das tecnologias digitais desde 1945, é difícil não pensar que as sugestões apresentadas por Bush sinalizariam a evolução tecnológica que estava por vir e que, de certo modo, concretizou-se a partir da Internet e de novos ambientes digitais (BUCKLAND, 1992). Contudo, esta é uma perspectiva incompleta e não histórica. Conforme destacam Nyce e Kahn (1989), embora publicado em 1945, o artigo de Bush foi escrito em 1939 e não possuía uma relação direta com computadores digitais, que só então começavam a ser inventados, ou com os computadores analógicos, em que o próprio Bush trabalhara.

Importante destacar que em 1945 uma ampla gama de desenvolvimentos em tecnologia da informação, incluindo rádio, televisão, filmes, cartões perfurados, microfilmes, circuitos e sensores fotoelétricos, já estavam bem desenvolvidos. Além disso, escritos de Paul Otlet (1934), Suzanne Briet (1951) e outros pioneiros da então denominada ‘Documentação’ tornam evidente que considerar o ano de 1945 como o ano do nascimento da Ciência da Informação é historicamente inadequado.

É inegável que a partir do desenvolvimento e popularização da *Web* o campo da Recuperação da Informação ganhou maior destaque e visibilidade, pois com a intensificação da utilização das tecnologias digitais novas possibilidades foram incorporadas aos processos de produção, armazenamento, representação e recuperação de informações.

### 3 MODELOS CLÁSSICOS E PESQUISAS CONTEMPORÂNEAS EM RECUPERAÇÃO DA INFORMAÇÃO

Com o aumento exponencial da quantidade de informações disponibilizadas em ambientes digitais, tornou-se imperativo o desenvolvimento de modelos de recuperação que possibilitassem a recuperação dos conteúdos informacionais disponibilizados. Entre inúmeras propostas de modelos de recuperação de informação identificados na literatura, podem-se destacar os denominados ‘modelos clássicos’ de recuperação da informação: o Modelo Booleano, o Modelo Vetorial e o Modelo Probabilístico.

O Modelo Booleano está fundamentado na teoria dos conjuntos e na Álgebra de Boole e caracteriza-se como um dos modelos de recuperação mais utilizados, devido à simplicidade de seu formalismo e facilidade de implementação. No modelo booleano, o conteúdo informacional dos documentos é representado por um conjunto de termos de indexação e as buscas são formuladas por meio de uma expressão booleana composta por termos ligados através dos operadores lógicos (AND, OR e NOT). O resultado de uma busca é composto por um conjunto de documentos cuja representação satisfaz as restrições lógicas da expressão de busca.

De acordo com Souza (2006, p. 166), a principal desvantagem do modelo booleano está ligada ao fato de o modelo trabalhar o conceito de relevância de forma binária, ou seja, “os documentos são analisados sob o critério dualista relevante/não relevante, e não é criada nenhuma espécie de ordenação dos resultados que atendam às condições de consulta”. Assim, o resultado de uma busca booleana caracteriza-se por uma divisão do *corpus* documental em dois

subconjuntos, separando os documentos que atendem à expressão de busca daqueles que não atendem, não indicando nenhum tipo de ordenação ou classificação dos documentos (FERNEDA, 2012). Apesar das desvantagens apontadas, o modelo booleano ainda é o mais utilizado em bancos de dados relacionais e catálogos de bibliotecas.

O Modelo Vetorial fundamenta-se na comparação parcial entre os termos de indexação que representam um documento com os termos da expressão de busca apresentada pelo usuário (KURAMOTO, 2002). Para cada termo é associado um valor numérico, entre 0 e 1, que define a sua relevância em relação à expressão de busca. A homogeneidade das representações, tanto dos documentos como das expressões de busca, possibilita calcular as similaridades existentes entre um documento e uma expressão de busca ou entre dois documentos distintos.

Diferente do que ocorre no Modelo Booleano, que apresenta apenas uma dicotomia entre os documentos que atendem e aqueles que não atendem à busca realizada, o Modelo Vetorial possibilita a ordenação do resultado da busca, considerando o grau de similaridade de cada documento em relação à expressão de busca apresentada pelo usuário.

O modelo Probabilístico está fundamentado na teoria matemática das probabilidades. Como destaca Souza (2006), esse modelo supõe a existência de um conjunto ideal de documentos que atenda a cada uma das expressões de busca possíveis de serem realizadas. A partir do resultado da primeira busca, o usuário seleciona os documentos que considera mais relevantes, tais documentos selecionados são submetidos a uma nova consulta juntamente com a expressão de busca, favorecendo um processo de refinamento interativo da pesquisa, conhecido como *Relevance Feedback*. Conforme destaca Ferneda (2012, p. 52):

O processo de recuperação de informação é caracterizado por seu grau de incerteza no julgamento de relevância dos documentos em relação à expressão de busca. Assim sendo, é mais realístico pensar em uma probabilidade de relevância do que em uma pretensa relevância exata, como a utilizada nos modelos booleano e vetorial.

Apesar do grande avanço que os modelos clássicos de recuperação representaram para o campo da Recuperação da Informação, devido às novas demandas resultantes da popularização da Web, tornou-se necessário (re)pensar, sob diferentes perspectivas, os modelos de recuperação.

Ao longo dos últimos anos, verifica-se uma tendência de pesquisas relacionadas à exploração de aspectos semânticos dos conteúdos informacionais e utilização intensiva de processamento computacional no intuito de contribuir para o aperfeiçoamento dos Sistemas de Recuperação da Informação (SRI). Entre inúmeras tecnologias e abordagens presentes na literatura, podem-se agrupar os principais avanços no campo da Recuperação da Informação em três grupos principais: Redes Neurais Artificiais; Tecnologias Semânticas; Processamento de Linguagem Natural (PLN).

As Redes Neurais Artificiais (RNA) fazem uma analogia ao funcionamento do córtex cerebral humano e possibilitam várias aplicações relacionadas à recuperação da informação, a partir de processos baseados em *Machine Learning* (ML), aprendizado de máquina, podem processar informações incompletas e identificar distorções por meio da realização de inferências automatizadas.

O primeiro sistema baseado em RNA, batizado como *Perceptron*, foi desenvolvido em 1957 por Rosenblatt, porém, devido às limitações computacionais da época, sua utilização para

fins práticos foi comprometida, de modo que apenas com os recentes avanços das tecnologias digitais tal campo de estudo voltou a ganhar destaque (DEHAVEN, 2017).

Nas últimas décadas, técnicas de *Deep Learning* têm sido utilizadas para a representação e recuperação de informações no ambiente *Web*. *Deep Learning* é caracterizada como um conjunto de técnicas de ML que recorre à utilização de algoritmos para a modelagem de abstrações e de alto nível, utilizando um grafo profundo com várias camadas de processamento (AREL; ROSE; KARNOWSKI, 2010).

Contudo, apesar das potencialidades apresentadas pelas técnicas de *Deep Learning*, sua utilização ainda é restrita principalmente às corporações comerciais e projetos com grande financiamento, devido principalmente ao nível de complexidade e poder computacional que tais técnicas exigem.

As Tecnologias Semânticas apresentam-se como uma abordagem mais próxima às atividades convencionalmente desenvolvidas no âmbito da Ciência da Informação, caracterizam-se como linguagens computacionais que possibilitam enriquecer semanticamente os conteúdos informacionais, por meio da incorporação de metadados e informações contextualizadas, utilizando linguagens como: XML, RDF e OWL (RAMALHO; OUCHI, 2011).

Para Marchiori (1998), as marcações semânticas proporcionam uma nova miríade de possibilidades para o projeto de mecanismos de recuperação de informações. Alvite Díez et al. (2010) destacam que por meio da incorporação de tais tecnologias nos processos de representação torna-se possível automatizar a compatibilização de vocabulários controlados em Sistemas de Recuperação de Informação.

Entre os novos padrões de representação utilizados para compatibilização de vocabulários controlados no ambiente *Web*, destaca-se o *Simple Knowledge Organization System* – SKOS, um modelo de dados para a representação formal da estrutura básica e conteúdo de Sistemas de Organização do Conhecimento (RAMALHO, 2015).

Neste contexto, destaca-se a necessidade da utilização de vocabulários controlados para favorecer uma melhor representação e conseqüentemente uma recuperação mais eficiente. Tal constatação apresenta-se atualmente como um dos elementos dificultadores para a recuperação de informações na *Web*.

Devido ao aumento exponencial da quantidade de informações disponíveis em ambientes digitais, a utilização de palavras-chave e descritores apresenta-se, cada vez mais, como um fator determinante para a qualidade dos SRI. Contudo, há uma grande quantidade de documentos que não possuem palavras-chave designadas em sua estrutura.

Considerando a falta de parâmetros definitivos e a dificuldade de elaborar manualmente vocabulários em ambientes não controlados, uma das alternativas apresentadas para resolver tal situação é a utilização de PLN para automatização de determinadas tarefas (ERCAN; CICEKLI, 2007).

O PLN consiste na utilização de algoritmos para a busca de informações relevantes em um *corpus* documental constituído a partir de uma grande quantidade de documentos. Conforme destaca Chowdury (2003), PLN é um campo de pesquisa interdisciplinar e de aplicação que explora como os computadores podem ser usados para processar e manipular textos em linguagem natural para fazer coisas úteis.

Tecnologias PLN já estão presentes em uma variedade de aplicações relacionadas com a área de Ciência da Informação, possibilitam automatizar tarefas como: criação de resumos;

indexação; tradução; reconhecimento de entidades; extração de relacionamentos; recuperação de informações relevantes, entre outras.

Moreiro González (2011) apresenta algumas das principais dificuldades relacionadas à Organização e Recuperação de informações na *Web*, destacando como aspectos dificultadores: a sobrecarga de Informação, a ineficácia das palavras-chave, a falta de autoridade literária e/ou a falta de sistemas com tecnologias PLN.

Nesta perspectiva, observa-se que tais dificuldades são comuns ao que ocorre em ambientes controlados, como no caso dos periódicos científicos eletrônicos, conforme abordado na próxima seção.

#### 4 APLICAÇÕES TECNOLÓGICAS EM PERIÓDICOS CIENTÍFICOS

No contexto das publicações científicas, os periódicos eletrônicos atendem as crescentes demandas de acesso imediato e reutilização efetiva dos resultados da pesquisa e de atividades que são importantes para interpretação, avaliação e reutilização da atividade ou de parte dela.

Para Kuramoto (2008, p.863), as revistas científicas “[...] desempenham, desde seu aparecimento, papel primordial na constituição de uma ciência oficial, validada, identificada, verdadeiro capital científico de referência, sobre a qual se apoiam as pesquisas subsequentes”.

Segundo Marcondes et al. (2008, p. 22),

O estatuto da guarda, preservação e disponibilização da cultura, dos conhecimentos em geral, e em especial do conhecimento científico, tão caro para as bibliotecas enquanto instituições, vem passando por um desafio com o surgimento da *web*.

Um artigo científico é um trabalho acadêmico que apresenta resultados de pesquisa realizada de acordo com alguma metodologia científica e publicado em periódicos ou anais de conferências. Segundo Sales (2013), o artigo é considerado fonte de conhecimento na comunicação formal da ciência.

Tendo como fio condutor as tendências de acesso livre, iniciativas em âmbito nacional e internacional romperam barreiras econômicas impostas pelo sistema editorial disponibilizando resultados de pesquisas científicas financiadas com recursos públicos a partir da disponibilização de periódicos eletrônicos na Internet.

No Brasil, a iniciativa da disponibilização de periódicos partiu do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que no ano de 2003 customizou o *software Open Journal Systems (OJS)*, um *software* de gerenciamento e publicação de revistas eletrônicas, desenvolvido para a construção e gestão de uma publicação eletrônica, isto é, desenvolvido pelo *Public Knowledge Project (PKP)*, da *University of British Columbia*, e deu-lhe o nome de Sistema Eletrônico de Editoração de Revistas (SEER).

O IBICT, após a tradução do *software OJS*, disponibilizou na *Web* a revista *Ciência da Informação*, primeiro periódico a utilizar essa tecnologia, assim inicia-se o processo de distribuição do SEER a editores brasileiros para que publicassem revistas científicas de acesso livre na *Web*.

O OJS contribuiu significativamente no aumento dos periódicos científicos eletrônicos, e por este motivo torna-se importante o desenvolvimento de tecnologias para o aperfeiçoamento

dos processos de busca e recuperação. Ao encontro desta demanda surgem os estudos de controle de vocabulário a partir da compatibilização das palavras-chave. O uso deste metadado representativo dos artigos científicos eletrônicos advém da grande importância para a busca e recuperação, porém a visualização das palavras-chave é ineficiente. A ferramenta *VCPC Tools* executa os procedimentos para tornar as palavras-chave compatibilizadas com um vocabulário controlado e possível de visualização e recuperação dos artigos a elas vinculados.

Por outro lado, a necessidade de interoperabilidade entre os sistemas de organização do conhecimento e o gerenciamento do vocabulário controlado do periódico passa a ser uma demanda remanescente, neste contexto. O uso do TemaTres é alvo de estudo para essa aplicação, assim como a possibilidade de integração do TemaTres e a *VCPC Tools*, cujo objetivo é aperfeiçoar a execução de busca dos artigos científicos.

#### 4.1 OPEN JOURNAL SYSTEMS

*Open Journal System (OJS)* é um *software* livre que gerencia e publica periódicos científicos na Internet, sua operação é realizada pelo editor para que seja feita a publicação da revista.

A partir da versão 3.0, o IBCT recomenda utilizar apenas o nome original OJS. Dentre as principais funcionalidades do OJS destacam-se:

- Instalação e gerenciamento local ou através do Periódico em Nuvens, dando liberdade para os editores;
- Editores configuram os requisitos, seções, processo de revisão, etc.;
- **Submissão é feita *online* pelos autores;**
- Suporte à revisão cega e duplo cega por pares;
- Gerenciamento de conteúdo (site da revista);
- Indexação do conteúdo e mecanismo de busca;
- Notificação por e-mail e sistema de comentários para leitores;
- Sistema de ajuda online sensível ao contexto;
- Suporta Acesso aberto (**recomendado**) ou exclusivo para assinantes;
- Diversas funcionalidades extras através de *plug-ins*.

Como o artigo em tela tem seu foco de interesse na representação visando à recuperação da informação, ou seja, a recuperação dos artigos publicados pelos periódicos eletrônicos qualificados, nosso interesse está particularmente no item “Submissão é feita *online* pelos autores”, visto que é na submissão feita pelos autores, diretamente no sistema do periódico, que serão gerados os metadados e na possibilidade do uso de metadados ou linguagem de marcação para os atributos que compõem o resumo estruturado. Resumo estruturado representa a versão precisa, sintética e seletiva do texto do documento, é feito a partir dos elementos que tenham maior importância para estruturação dos atributos: Introdução, objetivo, metodologia, discussão e resultados.

Alves (2010, p.47) define metadados como:

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Logo, o termo “metadados” adquire uma maior amplitude semântica para denominar as novas formas de representação e descrição dos recursos informacionais, em relação à linguagem XML, Furgeri (2006, p. 238) faz a seguinte afirmação: “[...] a XML passou a permitir que a própria estrutura do documento se transformasse em metadados.”

Os autores Dempsey e Heery (1997) pontuam que os padrões de metadados podem ser entendidos a partir de três níveis ou tipologias, conforme Quadro 1 a seguir:

**Quadro 1:** Tipologia de formato de metadados

	<b>Banda Um</b>	<b>Banda Dois</b>	<b>Banda Três</b>
Características do registro	Formato simples	Formatos estruturados	Formatos ricos
	Padrão proprietário	Padrões emergentes	Padrões internacionais
	Indexação do texto completo	Estrutura em campos	Identificação detalhada (descrição em campos ou <i>tags</i> codificadas)
Formato dos registros	Lycos, Altavista, Yahoo etc.	Dublin Core, IAFA Templates, RCF 1807, SOIF, LDIF	MARC, TEI, CIMI, EAD, ICPSR

**Fonte:** Dempsey; Heery (1997, p. 08, tradução nossa).

De acordo com o quadro 1, das tipologias denominadas Banda um, Banda dois e Banda três, é possível verificar diferentes tipos de padrões de metadados, que variam de estruturas simples de descrição, estruturas intermediárias e estruturas complexas de descrição.

Para a *National Information Standards Organization* (NISO) (2017, p. 10), as informações de metadados sobre o conteúdo de um recurso que possibilita que esse conteúdo seja encontrado ou entendido estão no âmbito dos metadados descritivos e apresenta os seguintes tipos de metadados:

- **Metadados de descrição** - Para encontrar ou entender um recurso.
- **Metadados administrativos** - Metadados técnicos (para descodificar e renderizar arquivos e para o gerenciamento de arquivos a longo prazo); Metadados de preservação (Direitos de propriedade intelectual) e Metadados de direitos (associados ao conteúdo).
- **Metadados estruturais** - Relações de partes de recursos entre si.
- **Linguagens de marcação** - Integra metadados e bandeiras para outros Recursos semânticos no conteúdo.

A linguagem de marcação integra metadados e conjuntos de conteúdo, dessa forma marcações podem indicar uma característica importante do recurso.

Para um recurso textual, como é o caso dos subconjuntos utilizados nos itens dos resumos estruturados, pode se fazer uma marcação estrutural nos elementos como parágrafos e palavras com informações semânticas. Os metadados estruturados permitem que usuários pesquisem ou ainda naveguem para encontrar informações de seu interesse. As propriedades de metadados ajudam na identificação e compreensão do recurso.

A Interoperabilidade, ou troca efetiva de conteúdo entre sistemas, baseia-se em metadados que descrevem esse conteúdo para que os sistemas envolvidos possam efetivamente perfilar o material recebido e combiná-lo com suas estruturas internas, também suportam gerenciamento de objetos digitais, fornecendo as informações necessárias para processar

conteúdo digital adequadamente ou entregar a versão apropriada para satisfazer as necessidades do usuário (NISO, 2017).

Para exemplificar e complementar as informações, a NISO apresenta os tipos de metadados, suas propriedades e qual seu uso primário, conforme o Quadro 2.

**Quadro 2:** Metadados: Tipos, propriedades e uso

<b>Tipo de Metadados</b>	<b>Exemplo de propriedades</b>	<b>Uso Primário</b>
Metadados de descrição	Título Autor Assunto Gênero Data de publicação	Descoberta Exibição Interoperabilidade
Metadados técnicos	Tipo de arquivo Tamanho do arquivo Data / hora de criação Esquema de compressão	Interoperabilidade Gerenciamento de objetos digitais Preservação
Metadados de preservação	<i>Checksum</i> Evento de preservação	Interoperabilidade Gerenciamento de objetos digitais Preservação
Metadados de direito	Status do copyright Termos de licença Titular dos direitos	Interoperabilidade Gerenciamento de objetos digitais
Metadados estruturais	Sequência Lugar na hierarquia	Navegação
Linguagens de marcação	Parágrafo Título Lista Nome Data	Navegação Interoperabilidade

**Fonte:** NISO (2017)

Zeng (2010) afirma que as linguagens de marcações são consideradas um recurso complementar aos metadados, porém as descrições dos metadados por si só não criam uma coleção rica e granular. Assim a autora supracitada pondera que: “[...] se faz necessário a convergência, especialmente em áreas onde os metadados e os esforços de marcação se sobrepõe”. (ZENG, 2010, p. 189). Alves (2010) alega que a construção adequada e padronizada de metadados constitui a chave para a efetividade dos sistemas de informações em ambientes digitais.

Isto posto, pode-se verificar a importância dos metadados e das linguagens de marcação como instrumentos para a representação da informação no ambiente *Web*, isto é, os estudos sobre metadados e suas aplicações possibilitarão que a representação da informação seja feita de maneira estruturada, o que otimizará e facilitará a recuperação da informação. Esta aplicação de interoperabilidade é possível de visualização por meio dos processos de coleta de metadados da ferramenta VCPC Tools, onde se utiliza o protocolo OAI-PMH do OJS, a partir dessa possibilidade de troca de informações os sistemas podem ser complementos integrativos reciprocamente.

## 4.2 VCPC TOOLS

Diante do contexto informacional, as publicações eletrônicas de periódicos científicos, disponíveis na web, vêm aumentando consideravelmente no âmbito da Ciência da Informação e demais áreas especializadas. As discussões sobre a organização do conhecimento e sua representação na web são alvo de pesquisas e estudos da Biblioteconomia, como também da CI.

A representação do contexto dos documentos, de maneira que seja a mais fidedigna do conteúdo, sempre foi e será uma tarefa árdua e complexa, portanto torna-se um grande desafio para os profissionais da área de Informação. Como consequência, esta tarefa de representar o contexto de documentos tem reflexo direto nos usuários, que têm a expressa necessidade de recuperar conteúdos de qualidade e de acordo com suas expectativas de leitura.

No XVI Enancib, em 2015, Santos e Cervantes (2015) publicaram os resultados dos procedimentos teóricos e metodológicos para o desenvolvimento do controle do vocabulário de periódicos científicos eletrônicos, a partir da compatibilização de palavras-chave dos artigos científicos eletrônicos. A pesquisa foi de natureza exploratória e descritiva com abordagem mista qualitativa e quantitativa. Os dados foram coletados e analisados por meio da aplicação da ferramenta *VCPC Tools*, que foi devidamente testada e recebeu, em 2017, licença-patente. A ferramenta *VCPC Tools* foi idealizada para o controle de vocabulário em periódicos científicos eletrônicos de áreas especializadas e encontra-se implantada nas revistas da Ciência da Informação e da Comunicação da Universidade Estadual de Londrina.

A aplicação da *VCPC Tools* envolve etapas e ações como: 1) Instalar a ferramenta VCPC Tools (ações: Identificar o endereço do repositório digital do periódico e criar a base de dados no sistema gerenciador de banco de dados); 2) Coletar e armazenar os vocabulários controlados (ações: identificar os vocabulários controlados da área especializada, estruturar os dados dos termos dos vocabulários controlados, descrever em linguagem SQL os termos dos vocabulários controlados, importar os dados dos termos dos vocabulários controlados); 3) Compatibilizar as palavras-chave com os termos dos vocabulários controlados (ações: coletar e armazenar os metadados, verificar as palavras-chave que são idênticas com os termos dos vocabulários controlados, verificar as palavras-chave que remetem à igualdade dos termos do vocabulário controlado, vincular as palavras-chave a termos dos vocabulários controlados); e 4) Tratar intelectualmente as palavras-chave não compatibilizadas automaticamente (ações: verificar as palavras-chave que tenham termos correspondentes nos vocabulários controlados, inserir a palavra-chave como termo no vocabulário controlado da revista). Os principais autores referendados na pesquisa supracitada para o desenvolvimento da metodologia – *VCPC Tools*: Gil Leiva e Alonso Arroyo (2005) – na questão do controle de vocabulário e palavras-chave; ANSI/NISO. Z39.19 (2005) – norma para elaboração de vocabulário; Orengo e Huyck (2001), Orengo, Buriol e Coelho (2007) – no tema compatibilização de termos e análise léxica e semântica.

O controle de vocabulário exerce um papel essencial no momento de atribuir palavras-chave a um artigo prestes a ser publicado. Indispensável para o uso de buscadores, a identificação de expressões costuma ser efetuada de maneira arbitrária e sem critérios, com prejuízo não somente para o resultado da busca, como também para a precisão no uso das referências. O estudo realizado neste artigo é visto como uma contribuição da área da Ciência da Informação para com a epistemologia e a metodologia científica para as áreas especializadas.

A representação do conhecimento é uma dimensão inerente ao controle de vocabulário. Há periódicos científicos que utilizam tesouros eletrônicos para representação do conhecimento, e o fazem por meio de um instrumento documentário para padronização dos termos. Outros disponibilizam nuvem de *tags* para apresentar o conteúdo do periódico de forma hierarquizada.

O controle de vocabulário é utilizado para aumentar a eficácia dos sistemas de recuperação, dos sistemas de navegação na *Web* e de outros ambientes que precisam identificar como localizar o conteúdo desejado por meio de algum tipo de descrição utilizando uma linguagem. A finalidade do controle de vocabulário é alcançar consistência na descrição de objetos de conteúdo e facilitar a recuperação. A necessidade do controle de vocabulário é decorrente de duas características básicas da linguagem natural: 1) duplicidade de nomenclatura para um mesmo conceito – exemplo: SOC/Sistemas de Organização do Conhecimento; 2) duplicidade de conceito para uma mesma palavra – exemplo: manga (camisa), manga (fruta). Para a obtenção do vocabulário controlado são necessários três principais métodos: 1) definição do escopo dos termos; 2) uso do relacionamento equivalente para a ligação de termos com sinônimos e quase sinônimos; 3) distinção entre homógrafos. (AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

Neste contexto, as palavras-chave têm um papel importante para a recuperação, porém nem todos os periódicos têm dado a atenção necessária para este viés. Em muitos periódicos, as orientações/normas de atribuição das palavras-chave não estão registradas de forma explícita. Comumente são vistas, nas regras de submissão, orientações para os autores tomarem como base e executar a atribuição, da melhor forma, como indício para uma representação fidedigna de seu texto. Este é um fato recorrente que ainda é pouco explorado cientificamente (GONÇALVES, 2008). As palavras-chave ou *keywords* foram aplicadas inicialmente nos primeiros índices automatizados em bases de dados, anteriormente nomeadas de bibliografias especializadas. A palavra-chave é uma representação do conteúdo significativo de um texto e é requerida no momento da necessidade de recuperação, ou seja, na formulação da estratégia de busca (FUJITA, 2004).

Este é um desafio contemporâneo para a área da Ciência da Informação: representar a informação para aprimorar a recuperação e o acesso, por meio de sistemas de organização do conhecimento, neste estudo, enfatiza-se a aplicação de vocabulário controlado para a recuperação de conteúdo temático em periódicos científicos eletrônicos. Portanto o controle de vocabulário torna-se importante no processo de recuperação, que por sua vez depende da disponibilização destes instrumentos para tratar tematicamente os artigos. A partir dos resultados da aplicação do controle de vocabulário nas palavras-chave dos artigos científicos eletrônicos, o autor tem condições e suporte para atribuir com segurança os termos que representam o conteúdo de seu artigo de forma fidedigna.

Por outro lado, percebe-se a necessidade de ferramentas para exercer a função de gestão dos termos do vocabulário controlado, conseqüentemente a integração com a interface de busca nos periódicos científicos eletrônicos. Neste contexto, a ferramenta *VCPC Tools* vem sendo estudada para o desenvolvimento de soluções de integração com *software* gerenciador de vocabulários controlados, o TemaTres.

### 4.3 TEMATRES

A aplicação de um *software* para a construção e manutenção de vocabulários controlados - tesouro, bem como outras linguagens consideradas mais complexas, contribui para a minimização de esforços operacionais do profissional de informação e a recuperação em catálogos *online* com padrões descritivos e temáticos, os quais utilizam, na construção de registros bibliográficos e de autoridades, os metadados. Um dos principais requisitos em um *software* de gestão de tesouro são as funcionalidades de construção e a possibilidade de manter, em constante evolução, a linguagem. Outro aspecto importante é a interoperabilidade do vocabulário controlado em ambientes multilíngues proporcionadas por meio da aplicação do *software* (FUJITA et al., 2017).

Precisa-se pensar nos procedimentos que sistematizam todas as atividades fins para a construção e gestão deste vocabulário controlado. Nesta perspectiva de integração da gestão do tesouro em sistemas computacionais, inserem-se os sistemas utilizados para este fim. Fujita et al. (2017) apresentam uma análise da literatura que aponta para os sistemas MultiTes e o TemaTres e aborda aspectos relevantes desses sistemas, como os principais requisitos.

O TemaTres tem como principal característica, e mais relevante para efeito deste estudo, ser um *software* livre, na sequência, e prioritariamente, a existência dos requisitos básicos, como inserção de termos, construção de relacionamentos como USE, UP, TG, TE, TR, cadastro de diversos tipos de notas (escopo, aplicação, bibliográfica). A possibilidade de interoperabilidade que tornou possível realizar este estudo. O TemaTres foi desenvolvido por Diego Ferreyra, em 2004, e é uma ferramenta *Web* para a gestão e representação do conhecimento por meio de vocabulários controlados, tesouros, glossários, taxonomias e outros modelos de representação do conhecimento (TEMATRES, 2018).

Santos, Cervantes e Londero (2017) apresentam um novo modelo de controle de vocabulário controlado para periódicos científicos eletrônicos, a partir da aplicação da VCPC *Tools* como aporte para desenvolvimento de um vocabulário controlado para a área da Comunicação. A revista utilizada como estudo experimental foi a “Discursos Fotográficos”, vinculada ao Programa de Comunicação da Universidade Estadual de Londrina (UEL). Vale ressaltar que não existiam instrumentos para suprir o perfil de publicações da revista.

O processo envolve etapas como: identificar o endereço do repositório digital do periódico; desenvolver o projeto físico do banco de dados; elaborar a estrutura conceitual do domínio e subdomínio do periódico em estudo; levantar os instrumentos terminológicos do domínio e subdomínio; analisar as palavras-chave com os termos dos instrumentos terminológicos; verificar as palavras-chave que são idênticas entre si; verificar as palavras-chave que tenham termos correspondentes nos instrumentos terminológicos; classificar tematicamente as palavras-chave conforme a estrutura conceitual do periódico; inserir a palavra-chave como candidata a um termo no vocabulário controlado da revista. Todo o processo foi realizado inicialmente de maneira semiautomática com a VCPC *Tools*, ao final foi importado no TemaTres para implementar um modelo de interoperabilidade entre os sistemas.

Santos, Fujita e Moreira (2018) realizam um trabalho de importação no TemaTres dos registros de autoridades compatibilizados, sua disponibilização está nominada como Tesouro Unesp, além da navegabilidade no tesouro está integrado com as buscas das bases de dados da Unesp. Os procedimentos de importação foram realizados por meio de mecanismos

semiautomáticos, para tanto foi desenvolvida uma ferramenta para realizar a conversão do formato MARC21 no suporte XML para texto etiquetado em arquivo TXT para possibilitar a execução da importação do Tesouro Unesp no TemaTres. A disponibilização do Tesouro Unesp para consulta de seus termos é parte fundamental da Política de Indexação das Bibliotecas Universitárias da Unesp (FUJITA, 2016), com finalidade de aprimoramento da recuperação da informação mais precisa por seus usuários durante a busca por assuntos.

Santos, Cervantes e Fujita (2018) realizam outro experimento de importação do Tesouro Brasileiro em Ciência da Informação (TBCI) de Pinheiro e Ferrez (2014) no TemaTres. Realizaram-se diversos procedimentos manuais e automatizados para conversão do TBCI disponibilizado em PDF para arquivo somente texto e organizados, por meio de planilhas eletrônicas, os relacionamentos dos termos. Os relacionamentos foram concretizados por meio de vínculos dos identificadores criados para cada termo. Para efeito de validação dos relacionamentos, eles foram inseridos em uma base de dados relacional, posteriormente exportados em formato texto etiquetado para importação no TemaTres.

Diante dos aspectos em torno do controle de vocabulário e do vocabulário controlado para exercer a função de instrumento de representação e recuperação da informação, cabe destacar a importância da disponibilização e integração destes instrumentos com interfaces de busca. Neste sentido, a integração somente é possível devido à interoperabilidade dos sistemas, que em sua grande maioria não contempla esta possibilidade.

O viés aqui tratado tem como foco os periódicos científicos eletrônicos gerenciados pelo OJS que estão no caminho de melhorar a *performance* de seus processos de busca. Iniciativas de pesquisas estão em desenvolvimento, porém ainda a demanda é grande e carece de fortalecimento de estudos na área.

## 5 CONSIDERAÇÕES FINAIS

Em conformidade com o exposto, cabe salientar que os aportes teóricos e tecnológicos que sustentam a elaboração de instrumentos para a representação e a recuperação da informação, em ambiente *Web* estruturado, encontram-se compatíveis com o objetivos de orientar a representação e a recuperação da informação.

A necessidade destes instrumentos, em especial no estudo de periódicos científicos eletrônicos, torna-se de extrema importância, tanto para o pesquisador quanto para o usuário que necessita obter resultados expressivos para a sua necessidade de busca. As palavras-chave, conjunto de termos atribuídos aos artigos científicos pelos autores para representá-los, são de grande importância para o processo de recuperação. Este é o *corpus* utilizado como subsídios para o controle de vocabulário em periódicos científicos eletrônicos, mesmo em processos de compatibilizações com vocabulários controlados já consolidados da área ou o *start* de um novo vocabulário controlado. Desta forma, valendo-se da efetiva importância das palavras-chave, que em muitas situações não é dada, além da invisibilidade nos sistemas de busca.

Vale considerar que esta necessidade de instrumentos como vocabulário controlado para a representação e recuperação vem ao encontro da demanda por ferramentas auxiliares para controle do vocabulário em periódicos científicos eletrônicos, neste caso utilizando a *VCPC Tools*.

A automatização dos processos está relacionada com a principal finalidade da abordagem sistêmica que é a redução de esforços humanos, conseqüentemente o aumento de produtividade. Isso se torna um aporte tecnológico para o profissional da Ciência da Informação, que pode dedicar-se exclusivamente para os aspectos cognitivos da produção de vocabulário controlado. Por outro lado, esta mesma perspectiva sistêmica aponta para a necessidade de ter um gerenciador de vocabulário controlado, TemaTres, que por sua vez exige a necessidade de integrações sistêmicas, pesquisa esta que se encontra em andamento.

## REFERÊNCIAS

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 134f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

ALVITE DÍEZ, M. L.; et al. Propuesta de representación del tesouro EuroVoc en SKOS para su integración en sistemas de información jurídica. **Scire: Representación y Organización del Conocimiento**, n.16, v.2, p.47-51, 2010.

AMERICAN NATIONAL STANDARDS INSTITUTE - ANSI; NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. **Z39:19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005.

AREL, I.; ROSE, D. C.; KARNOWSKI, T. P. Deep Machine Learning - A New Frontier in Artificial Intelligence Research. **IEEE Computational Intelligence Magazine**, v.5, n.4, p. 13–18, 2010.

BRIET, S. **Qu'est-ce que la documentation?** Paris: Édit - Éditions Documentaires Industrielles et Techniques, 1951. 48 p. Disponível em: <<http://martinetl.free.fr/suzannebriet/questcequeladocumentation/briet.pdf>> Acesso em: set. 2018.

BUCKLAND, M. K. Information as thing. **Journal of the American Societ**, New York, v. 45, n. 5, p. 351-360, 1991.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1p. 101-108, 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: set. 2018.

CASTELLS, M. Inovação, liberdade e poder na era da utopia tecnológica. In: MORAES, Dênis de (Org.). **Sociedade midiaticizada**. Rio de Janeiro: Mauad X, 2006. p. 225-231.

CHOWDHURY, G. C. Natural Language Processing, **Annual Review of Information Science and Technology**, v. 37, p. 51-89, 2003.

DEHAVEN, V. R. **Machine learning**: future capabilities and their implications. Tucson: University of Arizona, 2017. Disponível em: <<https://arizona.openrepository.com/handle/10150/320027>> Acesso em: out 2018.

DEMPSEY, L.; HEERY, R. **A review of metadata**: a survey of current resource description formats. 1997. Disponível em: [http://www.ukoln.ac.uk/metadata/desire/overview/rev\\_ti.htm](http://www.ukoln.ac.uk/metadata/desire/overview/rev_ti.htm) Acesso em: 26 nov. 2017.

ERCAN, G.; CICEKLI, I. Using lexical chains for keyword extraction. **Information processing and management**, v. 43, n. 6, p. 1705-1714, 2007.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Ciência Moderna, 2012.

FURGERI, S. **Ensino didático de linguagem XML**. São Paulo: Érica, 2001.

FUJITA, M. S. L. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Revista Brasileira de Educação Especial**, Marília, v. 10, n. 3, p. 257-272, set./dez. 2004.

FUJITA, M.S.L. (Org.) **Política de indexação para bibliotecas**: elaboração, avaliação e implantação. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. Disponível em: <[http://ebooks.marilia.unesp.br/index.php/lab\\_editorial/catalog/view/96/273/310-1](http://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/view/96/273/310-1)>. Acesso em: 31 jul.2018.

FUJITA, M. S. L.; SANTOS, L. B. P. DOS; CRUZ, M. C. A.; MOREIRA, W. Avaliação das características do Tema 3 e Multites para o controle de autoridades nas bibliotecas universitárias. **Scire**, Zaragoza, v.23, n.2, p.71-81, jul.-dic. 2017.

GIL LEIVA, I.; ALONSO ARROYO, A. La relación entre las palabras clave aportadas por los autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. **Revista Española de Documentación Científica**, Madrid, v. 28, n. 1, 2005.

GONÇALVES, A. L. Uso de resumos e palavras-chave em ciências sociais: uma avaliação. **Encontros Bibli**, Florianópolis, v. 13, n. 26, p. 1-15, 2008.

JANAITE NETO, J.; FERNEDA, E. Ontologia como recurso de padronização terminológica no processo de recuperação de informação. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 30-45, jan./jun. 2016. Disponível em: <<http://www.periodicos.ufc.br/index.php/informacaoempauta/article/view/2967/2692>>. Acesso em: 15 nov. 2018.

KURAMOTO, H. Acesso livre à informação: novos desafios. **Liinc em Revista**, Rio de Janeiro, v. 4, n. 2, p. 154 -157, 2008. Disponível em: <<http://revista.ibict.br/liinc/article/viewFile/3171/2834>> Acesso em: 20 jun. 2018

KURAMOTO, H. Sintagmas nominais: uma nova proposta para a recuperação de informação. **DataGramaZero** - Revista de Ciência da Informação, Rio de Janeiro, v. 3, n. 1, fev. 2002.

LIMA, J. L. de O.; ALVARES, L. Organização e representação da informação e do conhecimento. In: ALVARES, L. (Org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4 Editores, 2012. 248 p. cap. 1, p. 21-48.

MAKOWIECKY, S. Representação: a palavra, a idéia e a coisa. **Caderno de Pesquisa Interdisciplinar em Ciências Humanas**. Florianópolis. v. 57, p. 1-30. 2003.

MARCHIORI, M. The limits of web metadata, and beyond. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 7, 1998. Computer networks and ISDN systems. **Proceedings...**, v. 30. p. 1-9, 1998.

MARCONDES, C. H. et al. Ontologias como novas bases de conhecimento científico. **Perspectiva em Ciências da Informação**, v. 13, n. 3, p 20- 30, set./dez. 2008. Disponível em:< [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362008000300003](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362008000300003) > Acesso em 12 out. 2018.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, jan. /abr. 2001. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652001000100008&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100008&lng=pt&nrm=iso) > Acesso em: 12 maio 2018.

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

MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador : EDUFBA, 2011.

MOTA, D. A. R. **Representação e recuperação de informação em acervos digitais nos contextos das web semântica e web pragmática**: um estudo crítico. São Paulo. USP, 2015. 128f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação da Escola de Comunicação e Arte da Universidade de São Paulo. São Paulo. 2015

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding metadata**: what is metadata, and what is it for? Baltimore: NISO, 2017. Disponível em: <https://www.niso.org/publications/understanding-metadata-2017> Acesso em: 13 set. 2018.

NHACUONGUE, J. A. **O campo da Ciência da Informação**: contribuições, desafios e perspectivas da mineração de dados para o conhecimento pós-moderno. Marília: UNESP, 2015. 194f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências, UNESP – Marília. 2015.

NOVELLINO, M. S. F. A linguagem como meio de representação ou de comunicação da informação. **Perspectivas em Ciência da Informação**, v. 3, n. 1, p. 137-146, jul./dez. 1998. Disponível: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/602/371>> Acesso em: 5 jun. 2018.

NYCE, J. M.; KAHN, P. Innovation, pragmatism, and technological continuity: Vannevar Bush's Memex. **Journal of the American Society for Information Science**, v.40, n.3, 1989.

ORENGO, V. M.; BURIOL, L.; COELHO, A. A study on the use of Stemming for Monolingual Ad-Hoc Portuguese Information Retrieval. In: PETERS, C. et al. **Evaluation of multilingual and multi-modal information retrieval**. Berlin: Springer, 2007. p. 91-98.

ORENGO, V. M.; HUYCK, C. R. A stemming algorithm for the Portuguese language. In: INTERNATIONAL SYMPOSIUM ON STRING PROCESSING AND INFORMATION RETRIEVAL (SPIRE), 8., 2001, Chile. **Anais...** Laguna de San Raphael, Chile, 2001. p. 183-193.

OTLET, P. **Traité de documentation**: Le livre sur Le livre, théorie et pratique. Bruxelles: Mundaneum, 1934.

PALMER, S. E. Fundamental aspects of cognitive representation. In: ROSH, E.; LLOYD, B. B. (Ed.). **Cognition and categorization**. Hillsdale: Erlbaum, 1978.

PESCHL, M. F. Representing Representations. **The Semiotic Review of books**. Canada v. 8, n. 3., 2002.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: IBICT, 2014.

RAMALHO, R. A. S.; OUCHI, M. T. Tecnologias semânticas: novas perspectivas para representação de recursos informacionais. **Inf. Inf.**, Londrina, v. 16, n. 3, p. 60-75, jan./jun. 2011.

SALES, D. P. **Critérios de avaliação da produção científica em ciências sociais aplicadas**: inquirindo as bases de dados. 2013. 115f. Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo. Disponível em: <[www.scielo.org/local/content/pdf/S.1.pdf](http://www.scielo.org/local/content/pdf/S.1.pdf)>. Acesso em: 3 jul. 2017.

SANTOS, J. C. F. dos; CERVANTES, B.M. N. Controle de vocabulário em periódicos científicos eletrônicos: proposta de compatibilização de palavras-chave. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais...** João Pessoa: UFPB; ANCIB, 2015.

SANTOS, J. C. F. dos; CERVANTES, B.M. N.; LONDERO, R. R. Controle de vocabulário: representação e acesso aos conteúdos temáticos de um periódico científico de comunicação. In: LOPES, M. I. V. de L.; *et al.* (Org.). CONGRESSO INTERNACIONAL IBERCOM (15. : 2017 : Lisboa) XV Congresso IBERCOM 2017 : comunicação, diversidade e tolerância. **Anais...** São Paulo: ECA-USP; Lisboa: FCH-UCP, 2018. 6072 p.

SANTOS, J. C. F. dos; CERVANTES, B.M. N.; FUJITA, M. S. L. Tesouro Eletrônico: importação no TemaTres e disponibilização na web. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais...** Londrina: UEL; ANCIB, 2018.

SANTOS, J. C. F. dos; FUJITA, M. S. L.; MOREIRA, W. Tesouro Unesp: Integração do registro de autoridade para o TemaTres. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais...** Londrina: UEL; ANCIB, 2018.

SANTOS, P. L. V. C. As tecnologias na formação do profissional da informação. In: Valentin, M. I. P. (Org.). **Profissionais da informação: formação, perfil e atuação profissional**. São Paulo: Polis, 2000, p 103 -116.

SANTOS, P. L. A. C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. **DataGramZero**: revista de Ciência da Informação. v. 10, n. 3, 2009.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, dez. 2006. p. 161-173. Disponível em: < <http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf> > Acesso em: maio 2018.

STATEMENT of International Cataloguing Principles. [S.l.]: IFLA, 2009. Disponível em: [https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-en.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf). Acesso em: 18 set. 2018.

TEMATRES. **TemaTres**: servidor de vocabulários controlados. 2018. Disponível em: <<http://r020.com.ar/tematres/manual/>>. Acesso em: 27 jul. 2018.

ZENG, M. L. Linguagem de marcação específica por domínio e metadados descritivos: funções para a descoberta de recursos científicos. **Encontros Biblio**, Florianópolis, n. esp. 2. sem. 2010. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000009528/240a84fd024a998e01f7188294eb119f>> Acesso em: 22 jun. 2018