



XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-8 – Informação e Tecnologia

A RECUPERAÇÃO DA INFORMAÇÃO EM REPOSITÓRIOS DIGITAIS: UMA ANÁLISE DO USO DE ONTOLOGIAS PARA EXPANSÃO DA REPRESENTAÇÃO DE DOCUMENTOS NA BRAPCI.

INFORMATION RETRIEVAL IN DIGITAL REPOSITORY: AN ANALYSIS OF THE USE OF ONTOLOGIES TO EXPAND THE REPRESENTATION OF DOCUMENTS IN THE BRAPCI.

Liliane Pereira Gomes. UFCA.

André Anderson Cavalcante Felipe. UFCA.

Denysson Mota. UFCA.

Modalidade: Resumo Expandido

Resumo: Apresenta a Brapci como objeto de pesquisa, com o objetivo de expandir a quantidade de atributos que representam cada documento presente na plataforma, corroborando com a melhora no processo de representação e relevância nos resultados de buscas. Adota a pesquisa de tipo exploratório com abordagem quantitativa, com o uso do método da análise bibliográfica. Os resultados identificaram ontologias que podem expandir a quantidade de atributos que representam cada documento na Brapci. Conclui-se que as ontologias escolhidas pelo estudo podem promover melhorias no processo de representação da informação dos artigos na BRAPCI.

Palavras-chave: BRAPCI. Ontologia. Representação de Informação.

Abstract: This work presents Brapci as a research object, with the objective of expanding the number of attributes that represent each document present on the platform, corroborating the improvement in the representation process and relevance in the search results. It adopts exploratory research with a quantitative approach, using the method of bibliographic analysis. The results identified ontologies that can expand the number of attributes that represent each document in Brapci. It is concluded that the ontologies chosen by the study can promote improvements in the process of representing information in the articles in BRAPCI.

Keywords: BRAPCI. Ontology. Information Representation.

1 INTRODUÇÃO

No cenário atual, a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci) é um repositório digital (RD) que concentra grande parte da produção científica brasileira em Ciência da Informação (CI), contando com (BUFREM et al, 2010; BRAPCI, 2020):



- a) artigos de três (3) eventos: Encontro Brasileiro de Bibliometria e Cientometria (EBBC), Encontro Nacional de Pesquisa em Ciência da Informação (ENACIB) e a *International Society for Knowledge Organization* (ISKO Brasil);
- b) sessenta e oito (68) periódicos nacionais, tanto impressos como eletrônicos;
- c) quatorze (14) revistas internacionais.

A Brapci foi desenvolvida pelo grupo de pesquisa E3PI, da Universidade Federal do Paraná (UFPR) que indexa grande parte das publicações em CI produzidas no Brasil através de varredura, coleta automática, processamento, revisão e incorporação de artigos na íntegra que são disponibilizados sob política de acesso aberto (GABRIEL JUNIOR, 2014).

Nesse contexto, é relevante levantar uma discussão sobre a Representação da Informação em ambientes digitais, como o RD, porque em parte fazem a mediação entre a informação produzida dentro de uma instituição e pesquisadores de diversas áreas e lugares, geograficamente distantes, e somente compreendendo como o processo de análise de discurso e os processos de significação do conhecimento possuem aspectos semânticos e pragmáticas na sua construção podemos refletir sobre os desafios de estabelecer garantias que de fato atendam com precisão as requisições de busca em sistemas como os repositórios.

Entretanto, vale considerar que, apesar dos esforços da área da Biblioteconomia e da CI na produção de teorias e propostas práticas com ênfase nos processos de recuperação da informação no contexto digital, ainda existem lacunas a serem verificadas e isso vale para os processos de representação de documentos em RD. Frente ao exposto, surge o seguinte questionamento: uso de ontologias pode otimizar a quantidade de atributos de documentos representados na Brapci, de modo a melhorar as ações de recuperação da informação, nos resultados de buscas realizados pelos usuários?

Ciente da importância de contribuir com o desenvolvimento do campo interdisciplinar da Recuperação da Informação, bem como da Biblioteconomia, esta pesquisa aponta sua investigação para a Brapci, enquanto como lócus de aplicação, tendo como objetivo geral: promover qualidade na recuperação da informação na Brapci, através do uso de ontologias para expandir a quantidade de atributos que representam cada documento presente nela, corroborando com a melhora no processo de representação e possivelmente com a relevância nos resultados de buscas realizadas pelos usuários. Como objetivos específicos, busca-se:

- a) Desenvolver um modelo de identificação de etiquetas ontológicas para representação



de documentos, com base em teorias sobre (descrição de dados, relevância, ontologias, contexto, metadados, web semântica, web pragmática e de elementos contextuais - intratextual ou extratextual);

- b) Estabelecer um levantamento de ontologias que possam contribuir para expansão da representação de documentos da Brapci;
- c) Aplicar o modelo nas ontologias escolhidas pela pesquisa para identificar quais delas podem ser etiquetas ontológicas que favoreça a expansão da representação de documentos da Brapci;

Este estudo busca contribuir com o desenvolvimento de estudos acerca da qualidade das representações nos Sistemas de Recuperação de Informação (SRIs), mais especificamente nos RDs, e que está intimamente ligada à função do repositório e o seu uso, considerando que a representação impacta na recuperação das informações, proporcionando êxito para as pesquisas das comunidades acadêmicas de todo mundo.

2 ONTOLOGIAS

Não há definição aceita com unanimidade quanto ao conceito de ontologias, mas considera-se que elas fornecem representações de um conjunto de conceitos, relações e as suas propriedades com a finalidade de diminuir ambiguidades, permitir o compartilhamento e o reuso, e proporcionar mediação entre agentes computacionais. Keet (2020) expõe 4 definições para o termo ontologia, utilizadas por pesquisadores que tem a pretensão de esclarecer as propriedades desses artefatos usados para representar conjuntos de dados e algumas críticas quanto nebulosidade dos conceitos usados nas definições.

Um grupo de pesquisadores em Gaithersburg, Maryland (EUA), em um evento (*Ontology Summit*¹) em 2007 reuniu comunidades que trabalhavam para a construção desses artefatos e em conjunto produziram um diagrama que representa em duas dimensões, semântica e pragmática, fazendo uso de palavras chave para expor tudo o que envolve a produção de uma ontologia de domínio.

Guimarães (2015, p. 61) considera que “Ontologias são especificações formais e explícitas de conceituações compartilhadas. Uma ontologia captura o conceito sobre alguma coisa, o corpo de conhecimento formalmente representado é baseado em uma conceituação

¹ http://ontolog.cim3.net/wiki/OntologySummit2007_Communicate.html



dos objetos do mundo real”.

As ontologias como instrumento de representação podem ser distinguidas conforme Mota (2015) apresenta em sua pesquisa em três categorias de ontologias (fundamentação, domínio ou híbrida). As Ontologias de Fundamentação “são sistemas de categorias filosoficamente bem fundamentados e independentes de domínio que têm sido utilizados com sucesso para melhorar a qualidade de linguagens de modelagem e modelos conceituais” (GUIZZARDI, FALBO; GUIZZARDI, 2008, p.244), devido as suas características elas são relevantes para analisar e refinar terminologias identificando conflitos e inconsistências.

As ontologias de domínio são compostas de estruturas de conceitos que representam realidades específicas de uma área conhecimento. E as ontologias híbridas “tem representações de diferentes conceitos, tanto gerais ou abstratos, como também de objetos concretos ou reais dentro de um domínio” (MOTA, 2015, p.65).

2.1 Métodos de Avaliação de Relevância

Os componentes de buscadores e SRIs lidam com os seguintes aspectos: representação das necessidades de informação submetidas aos sistemas em forma de consultas fazendo uso de palavras chave; representações dos conteúdos; e métodos de avaliação de relevância.

Analisando do ponto de vista tecnológico, as aplicações podem aprimorar todo esse processo fazendo uso de estratégias para a monitoração de usuário, identificação de relações de autoridades, índices de citação e/ou realimentação de relevância para a construção de sistemas mais eficientes. Um exemplo é o método 5H1H, utilizado para investigar a qualidade de dados em contexto de SRI's.



Quadro 1 – Descrição de Dados usando o método 5W1H

Fase	Descrição
Why (Por quê?)	Refere-se ao entendimento dos dados e o contexto em que eles foram criados.
Who (Quem?)	Trata-se da descrição dos metadados relacionados com quem coletou os dados ou quem patrocinou a coleta.
What (O quê?)	Requer uma descrição apurada dos dados, orientada de modo a identificar: seus contextos digitais, seus parâmetros/ detalhes variáveis, suas informações, seus conteúdos de arquivos, suas informações taxonômicas adicionais e suas formas de organização.
When (Quando?)	Relaciona-se ao detalhamento dos aspectos temporais, como período (ano, mês, dia), tempo, regularidade e intervalos da coleta de campo.
Where (Onde?)	Direciona-se ao detalhamento dos aspectos de cobertura geográfica e espacial relacionados com o país, a cidade, local, coordenadas espaciais (latitude e longitude) e condições meteorológicas (clima, temperatura).
How (Como?)	Estabelece as condições de criação dos dados. Os detalhes são muito importantes para as possibilidades de reuso e interpretação futura dos dados.

Fonte: OLIVEIRA (2017, p. 185).

Vale ressaltar que o processo de desenvolvimento e teste de SRIs desenvolvidos em laboratório são orientados a sistemas, sendo assim possuem limitações quanto a aplicações reais.

A dimensão central da teoria cognitiva é um ponto a ser considerado, visto que, o processamento da informação que ocorre em diferentes níveis, mas não engloba somente o aspecto humano pois durante a comunicação os atores geram sinais para as máquinas. Assim a cada estágio desse processamento as informações podem transformar as estruturas de conhecimento do indivíduo e o seu comportamento de busca se torna potencialmente dinâmico

Vale ressaltar que o processamento de informações contextuais é o cerne da web pragmática. Agregar nível semântico e atuar através de vocabulários de referência, no entanto está direcionada a uma semântica globalmente aceita, pois, sua finalidade maior é a interoperabilidade (ARMENGAUD, 2006). O contexto é “a situação concreta em que os atos de fala são emitidos, ou proferidos, o lugar, o tempo, a identidade dos falantes etc., tudo o que é preciso saber para entender e avaliar o que é dito” (ARMENGAUD, 2006, p. 13). Assim o estágio cognitivo do processamento de informação onde as questões pragmáticas de implicações de informação se concentram, são pensados a nível coletivo.

A intencionalidade, que caracteriza a necessidade de um indivíduo, pode ser influenciada por diversos fatores, as estratégias de RI, táticas e avaliações de relevância são afetadas pelos estágios de desempenho a tarefa, o que torna o ato de satisfazer a necessidade do usuário um ato ainda mais complexo (INGWERSEN; JÄRVELIN, 2005). Os autores



esclarecem que, em um processo de representação da informação, um algoritmo só age em nível léxico ou sintático, raramente em nível semântico e nunca a nível cognitivo pragmático. Sendo assim, “a consequência é que qualquer intencionalidade, significado, contexto implícito e potencial informativo subjacente à mensagem gerada e comunicada são perdidos imediatamente” (INGWERSEN; JÄRVELIN, 2005, p. 37, tradução nossa).

Os indivíduos são sujeitos fragmentados, percebe-se então que suas experiências biopsicossociais não permitem que sistemas computacionais quantifiquem suas expectativas e definam um grau de normatividade muito preciso.

A depender do objeto de informação existem elementos de representação visíveis apenas com uma análise minuciosa e ainda assim a depender ótica de quem representa, dos limites da própria linguagem utilizada e do domínio sobre o assunto abordado no documento.

3 METODOLOGIA

Esta pesquisa é classificada como de tipo exploratório, pois “objetiva dar uma explicação geral sobre determinado fato através da delimitação do estudo, levantamento bibliográfico, leitura e análise de documentos” (OLIVEIRA, 2010, p. 65).

Quanto aos procedimentos metodológicos, a pesquisa é do tipo bibliográfica, baseada em levantamento de material já analisado, publicado e consolidado permitindo traçar os estudos desenvolvidos em representação e recuperação da informação em repositórios digitais (GERHARDT; SILVEIRA, 2009), assim como a realização de estudo experimental, que é frequentemente utilizado nas ciências tecnológicas, conforme Prodavov e Freitas (2013).

O método de análise bibliográfica adotado pela pesquisa (BASTOS, 2009), descreve as formas de busca, os locais, os sites etc., utilizando para isto, as seguintes técnicas básicas de leitura: leitura prévia, leitura exploratória, leitura, analítica e leitura interpretativa.

3.1 Modelo de Análise da Pesquisa

Mediante as técnicas do método bibliográfico proposto por Bastos (2009), em especial, as técnicas de leitura analítica e a leitura interpretativa, foram estabelecidos os procedimentos para a construção de uma ferramenta para a identificação de etiquetas ontológicas para representação de documentos na Brapci, conforme quadro a seguir:



Quadro 2 – Modelo de identificação de etiquetas ontológicas para representação de documentos (MIEORD)

Categorias de análise				
Questão de competência	Metadados	Intencionalidade	Contexto textual/intratextual	Contexto extratextual
Why (Por quê?)				
Who (Quem?)				
What (O quê?)				
When (Quando?)				
Where (Onde?)				
How (Como?)				

Fonte: elaborado pelos autores (2022).

A base teórica para a construção do modelo do Quadro 2 foi o método 5W1H (OLIVEIRA, 2017), para a questão da competência e a compreensão de metadados; o conceito de intencionalidade (INGWERSEN; JÄRVELIN, 2005), o conceito de contexto textual, intratextual e extratextual (ARMENGAUD, 2006) e o conceito de ontologias (MOTA, 2015).

4 IDENTIFICAÇÃO DAS ONTOLOGIAS

Neste momento, buscou-se avaliar se as ontologias que constituem o corpus de análise são capazes de auxiliar na representação das etiquetas de representação de documentos atendidas pelo padrão Dublin Core utilizado pela Brapci.

No processo de escolha das ontologias, a primeira ação realizada foi o levantamento de ontologias que contassem com elementos relacionados à atividade científica, para a partir delas identificar a existência de atributos que complementassem a Brapci, assim como a menção a outras ontologias candidatas para análise. Os resultados obtidos foram os seguintes:

- a) *Ontobee*;
- b) *Ontology Lookup Service*;
- c) *Linked Open Vocabularies (LOV)*;
- d) *SPAR Ontologies*.

Dentre as ontologias verificadas, apenas a LOV apresentou um ecossistema de ontologias com potencial para corroborar com os anseios da pesquisa. Sendo assim, após a análise de um montante considerável de ontologias, foram escolhidas que possuísem ações que corroborassem com a qualidade da recuperação de documentos. Conforme o quadro a seguir.



Quadro 3 – Ontologias analisadas pela pesquisa

The Data Use Ontology (DUO)
The Evidence & Conclusion Ontology (ECO)
Agent Relationship Ontology (AgRelOn)
Bibliographic Ontology (BibO)
Citation Typing Ontology (CiTO) e Citation Counting and Context Characterisation Ontology (C4O)
Scientific Evidence and Provenance Information Ontology (SEPIO)
Scholarly Contributions and Roles Ontology (SCoRO)
Document Components Ontology (DoCO)

Fonte: elaborado pelos autores (2022).

Diante do processo de análise algo que chama a atenção é a grande quantidade de ontologias já desenvolvidas com a finalidade de representar e recuperar informações na Web Semântica principalmente na área de saúde. Apesar do número considerável de aplicações específicas foi possível encontrar ontologias já existentes que podem agregar novas etiquetas de representação aos documentos da área de ciências sociais aplicadas.

O processo de identificação dos objetivos das propriedades de ontologias, na fase de levantamento de ontologias adequadas aos requisitos da pesquisa foram de extrema importância para compreender de que forma elas cumprem com o objetivo para a qual foi construída, e foi possível identificar em alguns casos que faltam descrições dessas propriedades exigindo maior esforço de interpretação.

Ao aplicar o Modelo de identificação de etiquetas para representação de documentos na Brapci para identificar as etiquetas ontológicas existentes, foi possível ver com clareza quais etiquetas poderiam ser integradas ao sistema sem causar redundâncias e assim enriquecer a experiência do usuário na plataforma. Importante lembrar que, essa avaliação funcionou como um primeiro filtro para seguir para a próxima fase: o levantamento de ontologias que pudessem contribuir para expansão da representação.



Quadro 4 – Identificação das ontologias que podem ser aplicadas a Brapci

Categorias de análise				
Questão de competência	Metadados	Intencionalidade	Contexto textual intratextual	Contexto extratextual
Why (Por quê?)		Objetivo SCoRO.		
Who (Quem?)			Tradutor BiBO	Referências bibliográficas; C4O Tipo de contribuídor (agentes); (AgRelOn); SCoRO;
What (O quê?)	Revisão de BiBO Transcrição de BiBO Tradução de BiBO			Elementos discurso. DoCO
When (Quando?)				Apresentado em BiBO
Where (Onde?)		Política de acesso DUO		Apresentado em BiBO Baseado perto BiBO
How (Como?)		Política de uso DUO Característica do uso da citação CiTO	Objetos de pesquisa. SEPIO	Evidências e Método SEPIO;

Fonte: elaborado pelos autores (2022).

O modelo proposto evidenciou quais as questões de busca, o uso de propriedades das ontologias (AgRelOn e CiTO) podem atender. Ao identificar as ontologias que possuem a adaptabilidade para o padrão *Dublin Core Metadata Iniciniative* é possível observar quais implementações podem ser feitas sem exigir muitas alterações nas estruturas de dados, pensando na compatibilidade e no contexto de comunicação em que a base se encontra.

5 CONSIDERAÇÕES FINAIS

Pela observação dos aspectos analisados, conclui-se que, se for possível representar toda a construção mental do conteúdo do documento, fazendo uso de RDF e Ontologias além da interoperabilidade entre sistemas é possível identificar intencionalidade. Como já mencionado, os contextos de busca podem ser inúmeros, mas mensurando os mais prováveis seguindo protocolos da Web Semântica e da Web Pragmática unido a mais esforço de processamento é possível aumentar o nível de identificação de relevância.

Sabe-se que a relevância é projetada na mente do usuário e que o processo de busca é de caráter dinâmico, por isso a necessidade pode ser modificada conforma a execução da tarefa. No entanto, quanto mais informação contextual (intencional, interacional e funcional)



as máquinas podem se aproximar do processamento de informações colocado por Ingwersen e Jarvelin (2005), pesquisas nesse contexto irão proporcionar que possamos compreender melhor o conhecimento científico disponível e processá-lo com maior praticidade.

Todo o percurso teórico construído neste estudo foi articulado de forma que o leitor pudesse ter um panorama contextual sobre o que de fato engloba a recuperação de informação, sobretudo a informação contida em suportes concretos. Sendo esta uma pesquisa que se concentra em contribuir de forma concreta para solucionar problemas da sociedade contemporânea a partir da Biblioteconomia, foi necessário relembrar o histórico de contribuições para a organização da informação, os desafios que cercam a prática dos profissionais que estão em diversos espaços (de informação, cultura, formação de leitores e tantos) e que ao longo do tempo foram incentivo para construir produtos para orientar a prática profissional.

AGRADECIMENTOS

O pesquisador Prof. Dr. Denysson Axel Ribeiro Mota agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento do Projeto MCTI/CNPQ Nº 436309/2018-1 – Universal, sob sua coordenação, e à Universidade Federal do Cariri (UFCA), pela concessão de recursos e bolsas de iniciação científica e de inovação tecnológica, que contribuíram para os estudos e reuniões interinstitucionais que auxiliaram a realização desta pesquisa.

REFERÊNCIAS

ARMENGAUD, Françoise. **A pragmática**. São Paulo: Parábola, 2006.

BRAPCI - **Base de Dados em Ciência da Informação**. [Dados dispersos]. Disponível em: <https://www.brapci.inf.br>. Acesso em: 17 jun. 2022.

BASTOS, R. **Ciências Humanas e Complexidades**: Projetos, métodos e técnicas de pesquisa. 2. ed. Rio de Janeiro: E-papers, 2009.

BREITMAN, K. K.; LEITE, J. C.S.P. Ontology as a requirements engineering product. **IEEE International Requirements Engineering Conference 2003**, Monterey Bay, CA, USA, p. 309-319, 2003. Disponível em: <http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Ontology%20as%20a%20RE%20product.pdf>. Acesso em: 17 jun. 2022.



BUFREM, L. S.; COSTA, F. D. O.; GABRIEL JUNIOR, R. F.; PINTO, J. S. P. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, v. 15, n. 2, 2010.

GABRIEL JUNIOR, R. F. **Geração de indicadores de produção e citação científica em revistas de Ciência da Informação**: estudo aplicado à base de dados BRAPCI. 140 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2014. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/123338/000823623.pdf?sequence=1&isAllowed=y>. Acesso em: 17 jun. 2022.

GUIMARÃES, Rachel Cristina Mello. **Nomeação de elementos ontológicos para criação de ontologias**: uma proposta metodológica. 2015. 169 f., il. Tese (Doutorado em Ciência da Informação). Universidade de Brasília, Brasília, 2015.

GUIZZARDI, G., FALBO, R., GUIZZARDI, R. S. S. A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software. **Revista IEEE América Latina**, v. 6, n.3, p. 244-251, 2008.

INGWERSEN, P.; JARVELIN, K. **The Turn**: Integration of Information Seeking and Retrieval in Context. Nederland: Springer, 2005. *E-book*.

KEET, C. M. **An introduction to ontology engineering**. 2020, Disponível em: <https://people.cs.uct.ac.za/~mkeet/files/OEbook.pdf> Acesso em: 17 jun. 2022.

MOTA, D. A. R. **Representação e recuperação de informação em acervos digitais nos contextos da web semântica e web pragmática**: um estudo crítico. 2015. Tese (Doutorado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://teses.usp.br/teses/disponiveis/27/27151/tde-27012016-135403/pt-br.php>. Acesso em: 15 jan. 2020.

OLIVEIRA, M. M. **Como fazer pesquisas qualitativas**. 3. ed. Petrópolis, RJ: vozes, 2010.

PRODANOV, C. C.; FREITAS E. C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2 ed. Novo Hamburgo: Feevale, 2013.

OLIVEIRA, A. C. S. Adoção de padrões de metadados para repositórios de dados digitais na ciência aberta. In: VERCHIATO, F. et al. (Org.) **Repositórios digitais**: teoria e prática. Curitiba: EDUTFPR, 2017. *E-book*.

RAUTENBERG, S.; TODESCO, J. L.; STEIL, A. V. Uma Ontologia para Instrumentos da Gestão do Conhecimento e Agentes da Engenharia do Conhecimento. **Informação & Sociedade: Estudos**, João Pessoa, v.21, n.1, p. 111-128, jan./abr. 2011. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/7172> . Acesso em: 17 jun. 2022.



SCHOOP, M.; MOOR, A.; DIETZ, J. L.G. The pragmatic web: a manifesto. **Magazine Communications of the ACM**, New York, v. 49, n. 5, p. 75-76, maio, 2006.