

Da epistemologia à implementação tecnológica: um olhar para o enriquecimento semântico na modelagem de domínio

José Eduardo Santarem Segundo¹

Ana Cristina de Albuquerque²

Resumo:

Com o conhecimento que se encontra em uma realidade concreta são desenvolvidas representações do conhecimento com o intuito de torná-las adequadas ao contexto e aos propósitos em que estão alocadas. A modelagem de domínios do conhecimento parte justamente da compreensão dessa realidade, pela abstração, e determina como os conceitos irão se relacionar em uma estrutura lógica, entre si e com o domínio específico. O presente trabalho propõe um olhar sobre a caracterização epistemológica da modelagem de domínio até a materialização tecnológica, com o objetivo de apresentar o enriquecimento semântico como um elemento que contribui para o processo de modelagem de um domínio do conhecimento. A pesquisa tem viés bibliográfico e exploratório e apresenta as técnicas de enriquecimento semântico por meio de ligações e por meio do uso de inferências, que vêm da caracterização de axiomas e definição de regras de inferência que podem ampliar as possibilidades de formalização do domínio. As considerações demonstram que o processo de modelagem de domínio pode trazer contribuições profícuas com a utilização das técnicas apresentadas, e que essas colaboram efetivamente para especificar elementos que individualizam e caracterizam um domínio do conhecimento.

Palavras-chave: modelagem de domínio; enriquecimento semântico; formalização de domínio; ontologia; representação do conhecimento.

¹ Livre Docente em Informação e Tecnologia pela Universidade de São Paulo (USP). Doutor em Ciência da Informação pelo PPGCI (UNESP). Docente do PPGCI (UNESP) e do DEDIC (USP). santarem@usp.br. <http://lattes.cnpq.br/5562746387565465>. <https://orcid.org/0000-0003-3360-7872>

² Doutora em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI UEL). <http://lattes.cnpq.br/4752632781155101>. <https://orcid.org/0000-0003-3506-0479>. albuati@uel.br

From epistemology to technological implementation: a look at semantic enrichment in domain modeling

Abstract:

With the knowledge that is found in a concrete reality, representations of knowledge are developed in order to make them adequate to the context and purposes in which they are allocated. The modeling of knowledge domains starts precisely from the understanding of this reality, through abstraction, and determines how the concepts will relate in a logical structure, among themselves and with the specific domain. The present work proposes a look at the epistemological characterization of domain modeling until technological materialization, with the objective of presenting semantic enrichment as an element that contributes to the modeling process of a knowledge domain. The research has a bibliographic and exploratory bias and presents semantic enrichment techniques through links and through the use of inferences, which come from the characterization of axioms and definition of inference rules that can expand the possibilities of domain formalization. The considerations demonstrate that the domain modeling process can bring fruitful contributions with the use of the presented techniques, and that these effectively collaborate to specify elements that individualize and characterize a domain of knowledge.

Keywords: domain modeling; semantic enrichment; domain formalization; ontology; knowledge representation.

1 INTRODUÇÃO

Este texto se inicia esclarecendo que a área de Organização e Representação do Conhecimento abarca inúmeras possibilidades de pesquisas, intrinsecamente relacionadas à Ciência da Informação.

Os autores propõem, no presente trabalho, explorar um viés tecnológico, transversal à Ciência da Informação e que, por vezes, também é transversal aos estudos sobre Organização e Representação do Conhecimento. Portanto, trilha-se um caminho fundamentado na epistemologia e nas bases teórico-metodológicas da Organização e Representação do Conhecimento, mas também direcionado para uma implementação e abordagem tecnológica.

O viés da Organização e Representação do Conhecimento pressupõe o relacionamento conceitual proveniente de abordagens e processos como o epistemológico, terminológico, a categorização, a escolha do ambiente e dos dados e do método que será utilizado, para delimitar o escopo de um domínio. Assim, a

utilização e interação entre diferentes técnicas possibilita um olhar que direciona o encadeamento das atividades de modelagem de domínio para uma dimensão teórica, metodológica e aplicada, que coaduna com a relevância de atender às especificidades e necessidades de diferentes usuários através de Sistemas de Organização do Conhecimento.

O viés tecnológico pressupõe descrever artefatos, ou seja, elementos computacionais capazes de transcrever computacionalmente as definições teóricas que a área de Organização e Representação do Conhecimento possibilita. Isso se traduz em pensar na objetificação computacional, transformando os conceitos, proposições e metodologias em funcionalidades concretas.

Nessa linha de pensamento, a modelagem de domínios caracteriza-se como um processo que parte da abstração para a definição de conceitos de determinado domínio do conhecimento, com o objetivo de consolidar um modelo estruturado e consistente na representação e na delimitação das relações conceituais.

Levando em consideração a transição entre o epistemológico e a materialização tecnológica, este trabalho apresenta o enriquecimento semântico como um dos elementos capazes de contribuir para a modelagem na formalização de um domínio.

O argumento principal é ter o enriquecimento semântico como elemento de caracterização, especificação e formalização de uma modelagem de domínio. O termo enriquecimento semântico, fundamental no nosso argumento, é fruto da chamada interoperabilidade semântica, que pode ser entendida como o elemento técnico capaz de gerar tal enriquecimento como resultado.

A perspectiva que se pretende discutir não é a técnica, e, portanto, optou-se pela escolha do termo enriquecimento semântico. Dito isso, a questão que pauta a interoperabilidade semântica está na construção e na viabilização da construção da relação semântica entre dois sistemas ou conjuntos de dados. Nesse contexto, a abordagem apresentada aqui tem natureza mais ilustrativa, para que o foco da discussão seja precisamente a capacidade de modelagem conceitual do domínio. Esse foco está embasado na premissa de que as condições técnicas e semânticas do processo técnico de interoperabilidade já estejam superadas.

Compreende-se que o enriquecimento semântico é a capacidade de ampliação da modelagem conceitual do domínio por meio de técnicas e recursos advindos de processos que incluem a interoperabilidade semântica, e que qualificam o modelo, seja do ponto de vista estrutural (metadados) ou dos dados.

A metodologia utilizada é de natureza bibliográfica e exploratória, com ênfase na caracterização do processo de modelagem de domínio. O enriquecimento semântico é apresentado a partir de duas técnicas amplamente utilizadas: o enriquecimento por meio de ligações juntando recursos de fontes diferentes por meio de um conector semântico; e por meio do uso de inferências, que parte da caracterização de axiomas e definição de regras de inferência que possam ampliar as possibilidades de formalização do domínio.

Considera-se que as técnicas aqui apresentadas podem fazer parte do processo de modelagem de domínios através da estruturação de ontologias, contribuindo efetivamente para especificar os principais elementos que caracterizam e individualizam um domínio.

2 MODELAGEM DE DOMÍNIOS DO CONHECIMENTO

Ao discutir e traçar paralelos entre as visões de mundo que constroem o conhecimento ocidental e oriental, Capra (1983) apresenta bases relacionais para o desenvolvimento do conhecimento científico. Para compreender um fenômeno, são elaboradas associações representando a realidade parcialmente e explicando aspectos da natureza, observados e testados por métodos científicos. A racionalidade do conhecimento científico advém das experiências empíricas e das observações do cotidiano. Essas pertencem à esfera do intelectual que discrimina, divide, compara, mede e categoriza, originando um “mundo” de atributos que se perfazem do contrário, mas que mantém relações entre si, por pertencerem a um sistema orgânico de saber (CAPRA, 1983).

A abstração, nesse sentido, é parte essencial do conhecimento científico, pois, a partir dela, a classificação e a comparação de formas, estruturas e fenômenos permitem emergir características específicas para construção de um “mapa intelectual” da realidade originário de um sistema de conceitos simbólicos,

abstratos e lineares que não representam completamente a realidade, mas que servem para comunicar a abrangência das estruturas que têm proximidade a esta realidade (CAPRA, 1983).

Esta compreensão aproximada pode ser feita por meio de modelos elaborados para representar uma parte específica de um fenômeno empírico ou um determinado domínio do conhecimento. Nas atividades científicas, os modelos têm um papel de “fôrma” onde se enquadram aspectos complexos que demandam investigação para compreensão de partes constituintes do todo.

Na literatura científica, os modelos têm definições de acordo com os diferentes ramos do conhecimento, conforme afirmam Sánchez, Caveró e Marcos (2005). Similarmente, os autores explicam que a criação de um modelo envolve o descarte e o agrupamento de características auxiliares na arquitetura da ideia de uma determinada realidade como também compreendê-la. Portanto, tem-se uma construção mental por meio da abstração e uma aplicação linguística referente à abstração, que é o modelo.

Dessa forma, os autores apresentam exemplos de definições genéricas como a da filosofia, que considera um modelo como a simplificação da realidade com a materialização da abstração; a da Matemática, que utiliza a Teoria dos Modelos e apresenta a relação entre linguagem formal, a Lógica, com sentenças e predicados para representar uma visão uniforme da realidade; a definição da Biologia, que concebe um sistema como a centralidade de seus estudos e tem o modelo como uma forma de comunicar os elementos deste sistema e como se relacionam entre si; a Ciência da Computação, em que os modelos são utilizados não só para representar algo, um modelo de dados, mas também como ferramenta com modelos relacionais (SÁNCHEZ; CAVERO; MARCOS, 2005). Consideram que as disciplinas compreendem o modelo como uma forma de compilar os principais atributos de algo por um processo de abstração de um domínio específico concretizado pela estruturação, descrição e representação (SÁNCHEZ; CAVERO; MARCOS, 2005).

No contexto dos sistemas de informação automatizados, Mendonça e Almeida (2012) explicam que os modelos podem ser definidos a partir dos elementos que possuem e de qual função eles cumprem. Quanto aos elementos, eles podem possuir

descrições, estruturas, teorias ou equações de determinado domínio do conhecimento; quanto à função, podem ser compreendidos como instrumentos representativos de uma parte ou objeto específico do mundo, teorias ou partes de teorias de domínios do conhecimento. De acordo com Torres, Almeida e Simões (2017, p. 03) a “[...] atividade de criação de modelos é realizada por indivíduos ou grupos nas diversas áreas do conhecimento e integra as raízes do método científico, já que todas as teorias são aproximações da verdadeira natureza das coisas”.

No âmbito da Organização e Representação do Conhecimento, a produção, disseminação e reprodução do conhecimento gerado por atores sociais dão origem a *constructos* que reúnem as representações inteligíveis e simplificadas do mundo. Tais *constructos* se encadeiam e demonstram características de um campo ou domínio do conhecimento (DODEBEI, 2002). A autora explica que essas representações podem ser estruturadas em modelos, com o objetivo de tornar a realidade complexa compreensível, ou seja, diante da definição das principais características, os modelos generalizam a realidade a partir de elementos relevantes, fundamentais ou interessantes, que, por sua vez, individualizam o objeto em seu domínio (DODEBEI, 2002).

Assim, simplicidade, estruturação e sugestibilidade são características de modelos que permitem a comunicação e disseminação de ideias com a possibilidade de serem replicados a outros grupos específicos, constituindo uma “[...] ponte entre os níveis da observação e do teórico.” (DODEBEI, 2002, p. 20). Os tipos de modelos são elaborados de acordo com a perspectiva de cada área do conhecimento, e para que as estruturas lógicas e conceituais sejam coerentes, deve-se reconhecer as especificidades do domínio do conhecimento em que se inserem.

Um domínio se caracteriza por sua especialização e possibilidade de delimitar, de acordo com uma comunidade de usuários, uma estrutura de conceitos que se relacionam e funcionam de modo a estabelecer um universo temático com autonomia terminológica (BARITÉ, 2000). Hjørland (2017) explica um domínio como um conjunto de conhecimentos, definidos social e teoricamente como o conhecimento de um grupo de pessoas que compartilham convenções ontológicas e epistemológicas. Os domínios podem ser disciplinas acadêmicas, mas também um

conjunto de atividades realizadas por uma comunidade discursiva. Diferentes teorias e interesses sociais podem constituir domínios de formas distintas. Essa diversidade torna complexa a análise do grau de consolidação do compartilhamento de informações em um domínio. É, portanto, fundamental buscar precisão nas visões teóricas e interesses subjacentes, pois sem uma estabilidade conceitual, um domínio pode não estar suficientemente sólido para ser estudado.

O processo de abstração, a análise e síntese resulta no delineamento de métodos, instrumentos e produtos, quando o acesso ao conhecimento registrado percorre um ciclo organizado e representado pelos Sistemas de Organização do Conhecimento (SOC). Os SOC são ferramentas semânticas com vocabulários estruturados e formalizados que permitem delimitar termos e conceitos específicos das áreas do conhecimento. Exemplos de SOC são vocabulários controlados, sistemas de classificação e ontologias, essas últimas, em especial, determinam os conceitos definidos por Hjórland (2017) como significados construídos de forma dinâmica, capazes de classificar o mundo por meio da intencionalidade com que foram idealizados. Portanto, não podem ser compreendidos de forma isolada, posto que há diferentes concepções e diferentes conceitos em cada domínio do conhecimento, demandando, assim, de um contexto específico. A ontologia: “[...] define os termos usados para descrever e representar uma área do conhecimento, pois podem ser usadas por pessoas, bancos de dados e aplicações que necessitam compartilhar informações em um domínio. (BRÄSCHER; CARLAN, 2010, p. 160).

Para Barité (2011), os SOC, de forma geral, incluem tanto as linguagens elaboradas para indexar e classificar quanto as linguagens que representam referenciais conceituais em estruturas semânticas relativas à documentação de qualquer área do conhecimento. O autor explica que o principal objetivo dessas linguagens é a representação temática de conteúdos de quaisquer tipos de documentos. Por isso, a opção ou delineamento de uma definição genérica nomeada como Sistemas de Organização do Conhecimento (BARITÉ, 2011) foi adotada, pois a diversidade apresentada norteia para uma nomenclatura mais ampla, que possa englobar as semelhanças e definir as particularidades.

Assim, os SOC têm a função de propiciar o uso social do conhecimento quando permitem a circulação da informação pelo relacionamento conceitual. Essa função pode ser percebida também a partir do direcionamento a serviços de referência, da utilização específica de conceitos em um domínio e como mediadores entre documento e usuário, conhecimento científico e documentação de classificadores e indexadores, no sentido de dar coerência e subsídios plausíveis para a construção de linguagens em ambientes controlados. (BARITÉ, 2011).

Os SOC são definidos como modelos da realidade que reúnem a complexidade da delimitação semântica e terminológica de domínios, assim como a determinação de métodos e etapas de elaboração que exigem uma modelagem do domínio de conhecimento. A atividade de modelagem é projetar, de acordo com um modelo, uma estrutura que reproduza ou descreva algo de forma simplificada, “[...] por meio de modelos criados a partir do processo cognitivo de realizar abstrações de recortes da realidade, tendo como objeto teorias ou fenômenos observados.” (TORRES; ALMEIDA; SIMÕES, 2017, p. 03). Deve ser elaborada segundo métodos e princípios norteadores que possibilitem autonomia quanto a representações e ao relacionamento conceitual que se modificam ao longo do tempo e dependendo do contexto em que estão inseridos (SÁNCHEZ; CAVERO; MARCOS, 2005).

Quanto aos métodos e às teorias, o estudo de Campos (2004) debate as perspectivas representacionais da Ciência da Informação, Terminologia e Ciência da Computação como bases que proporcionam um “núcleo comum” para a modelagem conceitual de um domínio do conhecimento. Da Ciência da Informação, a autora considera a Classificação Facetada de Ranganathan, que trabalha com a classificação do conhecimento a partir do método dedutivo e elege as categorias como fonte de ordenação de conceitos. Esses conceitos, por sua vez, originam as classes que demonstram o relacionamento conceitual em um domínio. Além disso, a autora considera a Teoria do Conceito desenvolvida por Dahlberg, para quem as classificações filosóficas seguem o princípio de que as categorias servem à compreensão de um conceito em um domínio do conhecimento. (CAMPOS, 2004). Da Ciência da Computação, a autora discorre sobre a modelagem de sistemas de bancos de dados com os modelos orientados a objetos e sobre a ontologia formal; e da Terminologia, as relações estabelecidas pelos conceitos em si e com o campo

semântico. Segundo a autora, essas abordagens podem auxiliar o processo de elaboração de modelagem conceitual e modelos de representação.

Neste sentido, Torres, Almeida e Simões (2017), identificaram duas concepções filosóficas que fundamentam teoricamente a elaboração do processo de modelagem: o Realismo, com a compreensão do conceito sob uma perspectiva universal, e o Conceitualismo, que apresenta o conceito como categorias, sem considerar universais. Os autores demonstram que, apesar de diferentes, as concepções se complementam por formularem classes hierárquicas e relacionamentos semânticos entre os conceitos. Além disso, as concepções favorecem delimitar teoricamente os processos de modelagem de domínios e a elaboração de modelos lógicos para a representação do conhecimento, como as ontologias.

2.1 Estruturação de ontologias e a etapa de modelagem de domínio

As ontologias possibilitam a visualização de um domínio hierárquico e os relacionamentos que se formam entre os conceitos, pois podem incluir a descrição e representação de objetos, ou parte deles, através de uma rede conceitual. (ALBUQUERQUE; CERVANTES, 2013). Para Barité (2011), as ontologias, além de caracterizadas como estruturas de conhecimento, são ilustrações de estruturas funcionais, com elementos e entidades que se relacionam e convergem em objetivos e funções determinadas (BARITÉ, 2011). O alicerce teórico e metodológico das ontologias como um SOC, está fundamentado no relacionamento entre as entidades e os assuntos, explicitando os elementos, as características e a rede conceitual do domínio (ALBUQUERQUE; HERNÁNDEZ QUINTANA, 2021). Dessa forma, Barité (2011, p. 132) explica que: “[...] Más que una estructura de conocimiento, una ontología es ante todo un sistema relacional de acciones que persigue tanto una gestión corporativa de calidad como la satisfacción plena del usuario”. Barité (2011) cita três níveis de conhecimentos que podem figurar em uma ontologia,

[...] a) el dominio, que contiene el conocimiento que constituye el núcleo del dominio, los conceptos y las relaciones permanentes entre ellos; b) el conocimiento que describe situaciones del mundo real. Las relaciones entre objetos son aplicadas a objetos o casos que son reales en un determinado momento, y no en otro, por eso son eventuales; y, c) el conjunto de deducciones e inferencias de se

extraen del dominio y de las situaciones del mundo real. (BARITÉ, 2011, p. 132).

Uma ontologia bem definida objetiva representar tanto os interesses dos usuários quanto seus comportamentos. Consequentemente, tem-se um domínio melhor definido justamente pela determinação de objetivos concretos relacionados aos aspectos fundamentais exibidos (BARITÉ, 2011).

As finalidades das ontologias são o reuso e o armazenamento do conhecimento e a interoperabilidade semântica e algumas tipologias e características são:

[...] se dividem em: **ontologias estáticas** que descrevem entidades existentes em um contexto de atributos e relações; **ontologias dinâmicas** que incorporam os aspectos dinâmicos dos contextos sociais e possuem foco nos estados e processos; **ontologias intencionais**, que expressam tarefas, objetivos e crenças de agentes situados em um determinado ambiente e as **ontologias sociais** que cobrem os aspectos sociais. (MOURA, 2011, p. 170).

Assim, Albuquerque e Hernandez Quintana (2021) explicam que as ontologias são instrumentos criados com o objetivo de organizar conceitualmente o conhecimento de um domínio específico para a representação de informações que contemplem a recuperação com coerência acerca do contexto em que será elaborada e empregada. Entende-se que os principais elementos característicos das ontologias são: possui um relacionamento conceitual em relação a um domínio; é estruturada em classes; possui um vocabulário formal e complexo em relação ao controle e à representação coerente do domínio; possui garantia literária; e permite o uso social do conhecimento (ALBUQUERQUE; HERNANDEZ QUINTANA, 2021). Portanto, domínios diferentes exigem metodologias diferentes, e esse fator influencia sua elaboração por ter um alto grau de complexidade para refletir as relações conceituais. Santarem Segundo e Coneglian (2015, p. 226) afirma que:

Utilizar ontologias é uma das maneiras de se construir uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação pelas ferramentas de recuperação da informação.

Para a estruturação de uma ontologia, pode-se optar por diferentes metodologias indicadas em algumas áreas do conhecimento, mas que apresentam um caminho

que tem em comum a modelagem de domínio. Rautenberg, Todesco e Steil (2008), fazem um levantamento dessas metodologias e apresentam algumas etapas fundamentais para a proposição de um modelo: etapa da especificação, a identificação do porquê e do objetivo de construção da ontologia; etapa de conceitualização: modelo conceitual; formalização: quando o modelo passa a ser formal; implementação: representa o conhecimento; e manutenção: revê a ontologia e a atualiza.

Para a modelagem de domínio, as etapas podem ser descritas como: a) seleção de conceitos; b) definição de conceitos; e c) estabelecimento de relações semânticas. Essas etapas incluem as teorias que fundamentam a modelização (CAMPOS, 2004; TORRES; ALMEIDA; SIMÕES, 2017); os princípios metodológicos que podem ser empregados (RAUTENBERG; TODESCO; STEIL, 2008; TORRES; ALMEIDA; SIMÕES, 2017) que vai derivar o relacionamento semântico do conhecimento.

Assim, considera-se relevante refletir a respeito do processo de modelagem conceitual de domínio do ponto de vista da aplicação tecnológica. Como indicado no início do texto, é o pensar sobre a possibilidade de estruturar artefatos que materializem computacionalmente a teoria que reflete sobre modelagem conceitual de domínio. Ou seja, quais as tecnologias, derivadas dos fundamentos da Organização e Representação do Conhecimento, permitem implementações possíveis de representar elementos diretamente atuantes na modelagem conceitual de um domínio.

3 ENRIQUECIMENTO DA MODELAGEM A PARTIR DE ONTOLOGIAS

A implementação tecnológica é uma capacidade computacional que representa uma proposição conceitual. É possível pensar em várias possibilidades de implementações, entretanto, aqui faz-se a opção pela discussão e compreensão de como o enriquecimento semântico pode ser entendido como um elemento importante para a modelagem conceitual de domínio.

Ressalta-se que o enriquecimento semântico, para o contexto abordado, parte da ideia da modelagem conceitual do domínio estruturada pelo uso de uma ontologia.

Conceitualmente, as ontologias desempenham um papel importante no processo de modelagem de domínio, fornecendo estruturação formal para representação de conceitos, das relações e das restrições em um determinado contexto. Elas permitem a organização e a estruturação do conhecimento, facilitando a compreensão e a comunicação entre diferentes partes interessadas.

Ao considerar as ontologias na modelagem de domínio, é possível capturar e representar de forma precisa as entidades e os relacionamentos relevantes, bem como suas propriedades e restrições. Isso permite uma descrição semântica comum, que facilita a interoperabilidade e o compartilhamento de conhecimento entre sistemas e domínios.

Para Guarino (1998, p.7, tradução nossa), ontologia é “uma maneira de se conceitualizar de forma explícita e formal os conceitos e restrições relacionados a um domínio de interesse”. Gruber (1993, p.2, tradução nossa) define ontologias como uma “[...] especificação explícita de uma conceituação.” Uma conceituação pode ser representada como um conjunto de objetos, restrições, relacionamentos e entidades que se assumem necessárias em alguma área de aplicação.

Para que se tornem efetivamente computacionais, as ontologias precisam passar de uma estrutura conceitual para implementação através de uma linguagem. Há um conjunto de linguagens desenvolvidas com esse propósito, entre elas a mais popular é a linguagem OWL, uma linguagem de marcação semântica para a definição, a instanciação, a publicação e a partilha de ontologias na Web. A linguagem OWL é reconhecida, atualmente, como o último padrão em linguagens para ontologia e é recomendada e considerada pelo consórcio W3C como a principal linguagem para construção de ontologias.

No entanto, somente uma linguagem não é o suficiente para transformar uma ontologia formal em um artefato tecnológico. Mesmo que ela permita uma materialização computacional para representar efetivamente todo o conceito que a ontologia propõe, são necessárias diretrizes como as descritas a seguir:

- O uso da linguagem para representação formal, como a OWL por exemplo, é o início do processo, permitindo que a ontologia possa ser interpretada por máquinas.
- A determinação de um ambiente ou framework que suporte a linguagem utilizada, para que possa ser feita a implementação da lógica.
- O processo de mapear classes e propriedades que caracterizam a estrutura da ontologia, incluindo a capacidade de relações com outros domínios do conhecimento.
- A construção de regras de inferência, de forma a incorporar um sistema de raciocínio no artefato computacional.
- A estruturação de uma estrutura computacional que permita a integração com sistemas e outras aplicações computacionais.

Entretanto, espera-se que para atender a essas diretrizes mínimas utilize-se algum tipo de formalização, geralmente encontradas nas metodologias de construção de ontologias. A adoção de uma metodologia ajuda a garantir a qualidade, a reusabilidade e a interoperabilidade das ontologias, promovendo assim a efetividade e o sucesso dos sistemas baseados em conhecimento. Entre as várias metodologias mais conhecidas destacam-se: Metodologia NeOn, Methontology e a Ontology Development 101.

Santarem Segundo e Coneglian (2015, p. 227) indicam que:

Para o uso como tecnologia da Web Semântica, entende-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias.

Ao considerar as ontologias como sistemas capazes de realizar modelagem de domínio, tende-se a relacionar a Organização e Representação do Conhecimento com todo o arcabouço teórico e tecnológico existentes na construção da Web Semântica, com seus primeiros passos no início dos anos 2000 com Tim Berners-Lee. É impossível descorrelacionar a Web Semântica de Berners-Lee aos modelos completos e complexos de sistemas de organização e representação do conhecimento que as ontologias podem nos proporcionar.

O objetivo deste estudo não é discutir ou avançar com os estudos sobre as ontologias, tão menos sobre Web Semântica, entretanto, considera-se esse um ponto de partida para a proposta principal do texto, qual seja, o processo de pensar a reestruturação de um domínio a partir das possibilidades que as ontologias podem proporcionar.

A partir desse cenário, segue-se para o foco principal da discussão. Apresenta-se duas possibilidades conceituais e tecnologicamente aplicáveis de ampliação da estrutura na modelagem do domínio a partir de ontologias. Uma delas ocorre por meio de conexões semânticas e a outra pelo uso de inferências.

3.1 Enriquecimento Semântico unindo domínios distintos

O enriquecimento semântico e a modelagem conceitual de domínio são duas abordagens complementares que objetivam melhorar os processos de Organização e Representação do Conhecimento em um determinado domínio/contexto. Ao combinar essas práticas, é possível potencializar a expressividade e a capacidade de inferência dos modelos conceituais.

Ao processo de adicionar informações semânticas a um modelo ou conjunto de dados, permitindo uma interpretação mais precisa e automatizada, também se atribui o termo enriquecimento semântico. O enriquecimento semântico é alcançado através do uso de ontologias, vocabulários controlados e anotações semânticas que fornecem uma estrutura conceitual comum e enriquecem o conteúdo com significado adicional.

As etapas necessárias ao enriquecimento dos dados são explicitadas de forma a apoiar o leitor na práxis de modelagem, utilizando-as para o enriquecimento semântico, ontologias e vocabulário controlado. Além disso, deve-se ainda observar as recomendações de boas práticas para a produção e disponibilização de dados abertos no contexto do *linked open data* (W3C, 2017).

Entende-se que discutir a ampliação do conjunto de dados é uma consequência requerida para o processo de enriquecimento semântico, entretanto, o enriquecimento semântico é um processo de qualificação da modelagem semântica do domínio. É possível pensar que o enriquecimento semântico é um processo que

nasce da necessidade de qualificar o conjunto de dados, mas que interfere e impacta diretamente a estrutura da modelagem.

Ao combinar o enriquecimento semântico com a modelagem conceitual de domínio, pode-se obter uma representação de conhecimento mais rica e estruturada. O enriquecimento semântico adiciona camadas de significado às entidades e aos relacionamentos, enquanto a modelagem conceitual estabelece uma base sólida para a organização e a visualização desse conhecimento. Na construção da modelagem conceitual do domínio, há a necessidade de abstrair e mergulhar nos conjuntos de dados para conceber a qualificação da estrutura.

Pode-se compreender o enriquecimento semântico como um plus na modelagem, a partir da possibilidade de conjuntos de datasets específicos para caracterizar o enriquecimento. Tecnicamente, a caracterização do enriquecimento semântico ocorre pela possibilidade (técnica) de conectar dois ou mais conjuntos de dados de domínios distintos por meio de um vocabulário semântico.

É difícil desmaterializar o enriquecimento semântico, porque apesar de alterar o pensamento estrutural do domínio, é somente por meio da relação entre conjuntos de dados que o enriquecimento semântico pode acontecer. Os conjuntos de dados que se conectam por meio de vocabulários semânticos são geralmente independentes entre si, com suas caracterizações específicas. É possível ilustrar (exemplificar) essa conexão entre os conjuntos de dados, que teoricamente deveriam, numa situação hipotética ideal, estar modelados sob uma ontologia de domínio.

Ao estruturar dois conjuntos de dados, como por exemplo, o conjunto de dados de publicações de um determinado periódico e o conjunto de dados do Currículo Lattes, é possível realizar uma conexão semântica entre esses conjuntos de dados por meio de um conector que deve ser um vocabulário semântico. Essa conexão pode acontecer por meio de um vocabulário que conecte, por exemplo, o ID do autor (nas duas bases) ou ainda o ID da produção (também nas duas bases), estabelecendo o link entre os conjuntos de dados.

A partir da interoperabilidade entre as bases por meio do conector, é possível juntar informações dos conjuntos de dados em consultas únicas a esses dados, relacionando entre si dados das duas bases. No exemplo apresentado, entende-se que há o enriquecimento ou qualificação das ontologias dos dois domínios.

Como mencionado anteriormente, não há como negar que é por meio da materialização computacional que se concebe o enriquecimento. Por mais que uma estrutura conceitual preveja a ligação entre as ontologias, é somente por meio de uma consulta que será possível desfrutar da ligação semântica existente entre dados das bases distintas. O uso de linguagens de ontologias e principalmente do protocolo SPARQL implementa computacionalmente o recurso e torna o enriquecimento um produto para consumo de usuários.

O SPARQL é um conjunto de especificações que fornece linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web. O padrão compreende as seguintes especificações: uma linguagem de consulta para RDF; uma especificação que define uma extensão do SPARQL para executar consultas distribuídas em diferentes terminais Query Language SPARQL; uma especificação que define a semântica de consultas SPARQL sob regimes de vinculação, como Schema, OWL, ou RIF; um protocolo RDF que define os meios para a transmissão de consultas SPARQL arbitrárias e as solicitações de atualização para um serviço de SPARQL; uma especificação que define um método de busca e descoberta e um vocabulário para descrever serviços SPARQL e um conjunto de testes, para avaliação da especificação SPARQL 1.1 (W3C, 2013).

3.2 Enriquecimento Semântico usando inferências

Uma das maneiras de qualificar um domínio, seja por meio dos metadados ou dos dados, envolve a construção de axiomas. Esses axiomas permitem que um sistema use a capacidade tecnológica de raciocínio e faça inferências em relação à uma modelagem de domínio. A inferência diz respeito à capacidade de se deduzir ou tomar decisões baseadas na consolidação de uma verdade de uma proposição que não é conhecida, mas é tida a partir da relação direta com outras verdades existentes. (SANTAREM SEGUNDO; CONEGLIAN, 2016).

É comum que sistemas baseados em ontologias usem raciocínio lógico para que possam executar inferências, enriquecendo o modelo e também o conjunto de dados aos quais são aplicados.

A inferência desempenha um papel fundamental na modelagem de domínio, permitindo a descoberta de conexões e o enriquecimento do conhecimento representado. Ao criar modelos de domínio, é crucial ir além das informações explícitas e explorar o contexto subjacente. Através da inferência, é possível extrair significados ocultos, identificar relações implícitas e preencher lacunas no conhecimento.

A inferência em modelagem de domínio se fundamenta na utilização de ontologias e regras lógicas para deduzir novas informações a partir das informações existentes. Isso pode ser alcançado por meio de técnicas como raciocínio lógico, raciocínio baseado em regras, raciocínio probabilístico e raciocínio baseado em caos.

Ao aplicar a inferência, é possível realizar consultas mais sofisticadas, obter respostas mais completas e tomar decisões mais informadas. Além disso, a inferência permite a detecção de inconsistências e a validação de modelos de domínio.

Em ontologias de domínio, os axiomas são declarações ou regras que estabelecem relações e restrições entre os conceitos e as propriedades definidas na ontologia. Eles descrevem as suposições fundamentais e as verdades lógicas consideradas válidas dentro do domínio específico da ontologia. Os axiomas são usados para estabelecer a estrutura conceitual e as relações semânticas entre os termos da ontologia.

Os axiomas ajudam a garantir a consistência, precisão e interpretação correta dos conceitos e das relações na ontologia. Eles fornecem um arcabouço lógico e formal para a representação do conhecimento em um domínio específico, permitindo inferências e raciocínio sobre os dados semânticos contidos na ontologia.

A implementação de axiomas e o uso de inferências só é possível a partir de implementações computacionais. Por mais que se possa pensar e estruturar

conceitos sobre uso de inferências, é a materialização que permite que elas efetivamente aconteçam.

Desta maneira, no contexto computacional, existem inúmeras formas matemáticas e algorítmicas para que as inferências sejam aplicadas e gerem a tomada de decisão, como inferências a partir de lógicas Fuzzy e bayesianas. A partir da ideia de serem realizadas inferências com o uso de dados e de determinadas regras, é necessário que existam motores de inferências para a geração das inferências em si. (SANTAREM SEGUNDO; CONEGLIAN, 2016).

Os motores de inferência são mecanismos capazes de atuar seguindo lógicas pré-definidas. Inicialmente, os motores de inferência foram criados dentro dos estudos de Inteligência Artificial, como uma ferramenta atuante em Sistemas Especialistas. Um Sistema Especialista pode ser entendido como uma aplicação criada para “[...] simular a ação de especialistas humanos, com o propósito de solucionar problemas específicos em um dado domínio.” (BRASIL, 1999, p. 3).

Nesse sentido, os motores de inferência são responsáveis por possibilitar que as expressões lógicas definidas possam gerar novos conhecimentos, visto que, esses mecanismos possibilitarão deduções e tomadas de decisões. Esses mecanismos necessitam dos axiomas para realizar a inferência, sendo que existem diversas linguagens para a criação das regras. (SANTAREM SEGUNDO; CONEGLIAN, 2016).

A implementação técnica depende da definição dos axiomas em linguagem específica. Atualmente a Semantic Web Rule Language (SWRL) é uma das linguagens recomendadas pelo W3C para essa função. A partir do momento em que os axiomas são escritos e o conjunto de dados é introduzido em um ambiente tecnológico com um motor de inferência, um Semantic Reasoner, o próprio ambiente realiza o processamento capaz de gerar a inferência, que resulta no enriquecimento, gerando alterações na modelagem do domínio.

Para ilustrar ou exemplificar um procedimento de enriquecimento semântico que altera a modelagem do domínio, é possível estruturar uma ontologia básica para um cadastro governamental, como o INSS ou algum outro similar. A estrutura organizacional do domínio prevê metadados (propriedades) para o titular e seus

respectivos pais (pai e mãe), em uma hierarquia vertical de apenas um grau. A partir desse conjunto de dados é possível escrever axiomas que permitam interpretar que quando A é pai de B, e B é pai de C, A é avô de C, uma informação hierárquica com dois graus de verticalidade, que inexistente na modelagem do domínio.

A partir desse exemplo é possível pensar nas implementações horizontais, baseadas em: se A é filho de B e C, e X é filho de B e C, então A é irmão de X. Ou seja, é possível criar inúmeras possibilidades de enriquecimento ao domínio.

Uma boa experiência na construção conceitual dos axiomas pode gerar enriquecimento de muita qualidade para um domínio, ampliações que nunca seriam possíveis sem o uso de técnicas computacionais.

Dessa forma, é importante ressaltar que pensar o enriquecimento semântico é repensar a organização da modelagem do domínio após ela já ter uma definição inicial. É um processo posterior à primeira definição da modelagem conceitual do domínio, porque ela parte de um pressuposto de um domínio já pronto e que deve ser ampliado a partir de inferências ou de ligações com outros domínios disponíveis.

4 CONSIDERAÇÕES FINAIS

Ao propor discutir o enriquecimento semântico como forma de caracterização do processo de modelagem de domínio, o presente trabalho o evidencia não só para as bases de fundamentação teórica, mas principalmente para a possibilidade de consolidar princípios de aplicação capazes de oferecer resultados coerentes na organização, representação e recuperação de informações.

Destacou-se a importância da implementação tecnológica para o uso de inferências e o enriquecimento semântico na modelagem conceitual de domínio. Ficou evidente que a incorporação de tecnologias semânticas e a aplicação de técnicas de inferência permitem enriquecer o significado dos dados, tornando-os mais estruturados e inteligíveis, conseqüentemente, ampliando as possibilidades de caracterização de um domínio.

A implementação tecnológica adequada aliada ao enriquecimento semântico possibilita a criação de ontologias robustas, capazes de representar com precisão o

conhecimento de um domínio específico. Desse modo, facilita a interoperabilidade entre dados de domínios distintos, a recuperação da informação e o desenvolvimento de serviços avançados.

Identificou-se também que o uso de inferências baseadas nas relações semânticas estabelecidas na ontologia permite que os sistemas realizem raciocínios lógicos e dedutivos, ampliando sua capacidade de análise e tomada de decisão. Isso é especialmente valioso em cenários complexos, nos quais a interpretação e a correlação de dados são cruciais.

Para concluir, a implementação tecnológica adequada para o uso de inferências e enriquecimento semântico na modelagem conceitual de domínio é fundamental para obtenção de sistemas mais inteligentes e eficientes. Essa abordagem impulsiona a evolução da Ciência da Informação, permitindo que a informação seja representada, compartilhada e utilizada de forma mais eficaz, contribuindo para a criação de ambientes informacionais avançados e aprimorando a tomada de decisões em diversos setores.

REFERÊNCIAS

ALBUQUERQUE, A. C.; CERVANTES, B. M. N. Ontologias a partir de parâmetros da norma ansi/niso z39.19. *In: ISKO BRASIL, 2., 2013, Rio de Janeiro. Anais [...].* Rio de Janeiro: ISKO Brasil, 2013 p. 177-182.

ALBUQUERQUE, A. C.; HERNANDEZ QUINTANA, A. R. Sistemas de organização do conhecimento e humanidades digitais: Possíveis interlocuções a partir da abordagem da análise do domínio. *In: SILVA, C. G.; REVEZ, J.; CORUJO, L. (org.). Organização do Conhecimento no Horizonte 2030: desenvolvimento sustentável e saúde.* Lisboa: Colibri, 2021. p. 727-732.

BARITÉ, M. *Diccionario de organización y representación del conocimiento: clasificación, indización, terminología.* Uruguai, 2000. Disponível em: <http://www.eubca.edu.uy/diccionario/index.htm>. Acesso em: 16 abr. 2023.

BARITÉ, M. Sistemas de Organização do Conhecimento: uma tipologia atualizada. **Informação & Informação**, Londrina, v. 16, n. 2, p. 122, 15 dez. 2011. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2011v16n2p122>. Acesso em: 31 mar. 2023.

BRASIL, L. M. Proposta de arquitetura para sistema especialista híbrido e a correspondente metodologia de aquisição do conhecimento. 1999. 256 f. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis. 1999. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/81196/137930.pdf?sequence=1>. Acesso em: 31 mar. 2023.

BRÄSCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: Jaime Robredo; Marisa Bräscher (Orgs.). *Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento*. Brasília DF: IBICT, 2010, p. 147-176. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso em: 31 mar. 2023.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, [s.l.], v. 33, n. 1, p. 22-32, abr. 2004. Disponível em: <http://dx.doi.org/10.1590/s0100-19652004000100003>. Acesso em: 13 jun. 2023.

CAPRA, F. *O tao da física*. São Paulo: Cultrix, 1983, 160 p.

DODEBEI, V. *Tesouro: linguagem de representação da memória documentária*. Rio de Janeiro: Interciência, 2002, 120 p.

GRUBER. *A translation approach to portable ontology specifications*. Stanford: Knowledge Systems Laboratory, 1993. 315 p.

GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 1998, Amsterdam. *Proceedings [...]* Amsterdam: IOS Press, 1998. p. 3-15.

HJØRLAND, Birger. Domain Analysis. *Knowledge Organization*, [s.l.], v. 44, n. 6, p. 436-464, 2017. Disponível em: <http://dx.doi.org/10.5771/0943-7444-2017-6-436>. Acesso em: 13 jun. 2023.

MENDONÇA, F. M.; ALMEIDA, M. B. Modelos e teorias para representação: uma teoria ontológica sobre o sangue humano. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: Ancib, 2012. p. 1-20. Disponível em: <https://brapci.inf.br/index.php/res/v/175173>. Acesso em: 13 jun. 2023.

MOURA, M. A. Interoperabilidade Semântica e Ontologia Semiótica: a construção e o compartilhamento de conceitos científicos em ambientes colaborativos online. *Informação & Informação*, [s.l.], v. 16, n. 2, p. 165, 15 dez. 2011. 2p165. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2011v16n>. Acesso em: 13 jun. 2023.

RAUTENBERG, S.; TODESCO, J. L.; STEIL, A. V. Uma metodologia para o desenvolvimento de ontologias. *Revista de Ciências Exatas e Naturais*, v. 10, n. 2, p. 237-262, 2008.

SÁNCHEZ, D. M.; CAVERO, J. M.; MARCOS, E. On models and ontologies. *In: WORKSHOP ON PHILOSOPHICAL FOUNDATIONS OF INFORMATION SYSTEMS ENGINEERING*, 1., 2005. [s.l.]. *Anais...2005*. p. 1-11.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Tecnologias da Web Semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. *In: ORGANIZAÇÃO DO CONHECIMENTO E DIVERSIDADE CULTURAL*, 3. Marília: Fundepe, 2015, p. 224-233. Disponível em: <http://isko-brasil.org.br/wpcontent/uploads/2015/09/Organiza%C3%A7%C3%A3o-do-Conhecimento-eDiversidade-Cultural-ISKO-BRASIL-2015.pdf>. Acesso em: 13 jun. 2023

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web Semântica e Ontologias: um estudo sobre construção de axiomas e uso de inferências. *Informação & Informação*, [s. l.], v. 21, n. 2, p. 217-244, 2016.. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/26417>. Acesso em: 13 jun. 2023

TORRES, S; ALMEIDA, M. B.; SIMÕES, M. G. M. Princípios para modelagem de domínios em sistemas de organização do conhecimento (SOC). *In: Congresso ISKO Espanha-Portugal*, 3., 2017, Coimbra, Portugal. *Anais [...]*. Coimbra: ISKO, 2017. p. 1-10.

W3C. *Data on the Web Best Practices*. 2017. Disponível em: <http://www.w3.org/Translations/DWBP-pt-BR/>. Acesso em: 13 jun. 2023.

W3C. *Sparql 1.1 Overview*. 2013. Disponível em: <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>. Acesso em: 13 jun. 2023.