

# **Modelando domínios: uma revisão sobre habilidades profissionais, práticas, metodologias, tecnologias e instrumentos de representação**

---

*Fernanda Farinelli*<sup>1</sup>

*Amanda Damasceno de Souza*<sup>2</sup>

*Marília de Abreu Martins de Paiva*<sup>3</sup>

*Fabiola Aparecida Vizentim*<sup>4</sup>

## **Resumo:**

A evolução das tecnologias da informação traz desafios à recuperação automática. Buscadores têm dificuldade em dar sentido à linguagem humana comprometendo taxas de revocação e precisão. Cresce a demanda pelo tratamento da informação atendendo requisitos como padronização e interoperabilidade. A atuação do profissional da informação assume uma perspectiva interdisciplinar e multifacetada, promovendo habilidades necessárias à representação do conhecimento visando construir vocabulários semânticos, sem ambiguidades, legíveis por humanos e máquinas. Na modelagem de domínio, a representação dá-se por um conjunto de termos e relações entre termos que traduzem os principais conceitos das “coisas” deste domínio. Na representação do domínio, utiliza-se diferentes instrumentos de representação, com diferentes graus de formalismo e complexidade, adotando técnicas e ferramentas na modelagem. A análise do domínio de conhecimento requer diferentes níveis e perspectivas de compreensão, exigindo diversas competências em informação do modelador na construção de artefatos de representação, deve-se ainda adotar técnicas de aquisição de conhecimento para compreensão do domínio. O artigo parte de uma abordagem conceitual fornecendo uma explanação didática dos conceitos necessários a sua compreensão, discute as contribuições dos campos da organização e representação

---

<sup>1</sup> Professora Adjunta (FCI/UnB). Doutora em Gestão e Organização do Conhecimento (PPGOC/ECI/UFMG). Ontologista responsável pelo projeto da OntONeo. fernanda.farinelli@unb.br. Lattes <http://lattes.cnpq.br/1907817850408525>. ORCID 0000-0003-2338-8872.

<sup>2</sup> Professora do Programa de Pós-graduação em Sistemas de Informação e Gestão do Conhecimento (FUMEC/MG). Doutora em Gestão e Organização do Conhecimento (PPGOC/ECI/UFMG). Bibliotecária Clínica. amanda.dsouza@fumec.br. Lattes <http://lattes.cnpq.br/3615797323442040>. ORCID 0000-0001-6859-4333.

<sup>3</sup> Professora Adjunta (DOTI/ECI/UFMG). Professora do Programa de Pós-graduação em Gestão e Organização do Conhecimento (ECI/UFMG). Doutora em Ciência da informação (PPGCI/ECI/UFMG). mariliapaiva@ufmg.br. Lattes: <http://lattes.cnpq.br/5084323621859190>. ORCID-ID: 0000-0002-0155-4043.

<sup>4</sup> Bibliotecária com experiência em dados, informação, conteúdo e gestão do conhecimento para negócios e tecnologia. fabiolavizentim@alumni.usp.br. Lattes <http://lattes.cnpq.br/8830340105267762>. ORCID [0000-0003-0181-085X](https://orcid.org/0000-0003-0181-085X)

da informação com enfoque na modelagem de domínio visando identificar as técnicas, metodologias, tecnologias e instrumentos de representação necessários para essa atividade. Como contribuições apresenta-se um conjunto de competências e habilidades essenciais para profissionais de informação modelarem domínios. Essa abordagem visa suprir lacunas técnicas não abordadas na formação tradicional destes profissionais, proporcionando maior compreensão dos aspectos técnicos relacionados à modelagem de domínio e permitindo ampliação de sua área de atuação.

**Palavras-chave:** modelagem de domínio; competências do profissional da informação; organização e representação da informação; profissional da informação; competência informacional.

*Modeling Domains: A Review of Professional Skills, Practices, Methodologies, Technologies, and Representation Tools.*

**Abstract:**

The evolution of information technologies brings challenges to automatic retrieval. Search engines struggle to make sense of human language, compromising recall and precision rates. There is a growing demand for information processing that meets requirements such as standardization and interoperability. The role of information professionals assumes an interdisciplinary and multifaceted perspective, promoting the necessary skills for knowledge representation aimed at building semantic vocabularies that are unambiguous and readable by humans and machines. In domain modeling, representation is achieved through a set of terms and relationships that translate the key concepts of "things" in this domain. Various representation tools with different degrees of formality and complexity are used in domain representation, employing techniques and tools in the modeling process. Analyzing a knowledge domain requires different levels and perspectives of understanding, demanding diverse information competencies from the modeler in constructing representation artifacts. Additionally, knowledge acquisition techniques are essential for a deep understanding of the domain. The article takes a conceptual approach, providing a didactic explanation of the necessary concepts and discussing the contributions of information organization and representation fields, focusing on domain modeling to identify the required techniques, methodologies, technologies, and representation tools for this activity. As contributions, a set of essential competencies and skills are presented for

information professionals to model domains. This approach aims to address technical gaps not covered in traditional training for these professionals, providing a better understanding of the technical aspects related to domain modeling and expanding their scope of practice.

**Keywords:** domain modeling; information professional competencies; information organization and representation; information professional; information competence.

## 1 INTRODUÇÃO

Com o aumento constante de dados gerados nas organizações diariamente, é cada vez mais importante que elas sejam capazes de gerenciar e utilizar esses dados de maneira eficiente e eficaz. Este aumento de dados é ao mesmo tempo impulsionado e manipulado por meio de diversas tecnologias da informação e comunicação (TICs) que promovem o processo de datificação, dentre as quais podemos citar a Web de Dados, Big Data, internet das coisas, dentre outras.

O termo "datificação" ou "dataficação" é usado para se referir ao processo de transformação de dados em formato digital para que possam ser armazenados, processados e analisados posteriormente (MAYER-SCHÖNBERGER; CUKIER, 2013). A internet das coisas (IoT, do inglês *Internet of Things*) é uma das principais fontes de dataficação na atualidade, pois gera enormes quantidades de dados de diferentes tipos e formatos.

A IoT é um conceito que se refere à conexão de objetos do cotidiano à internet, permitindo que eles possam ser monitorados, controlados e até mesmo tomar decisões de forma autônoma. O termo foi cunhado pelo pesquisador britânico Kevin Ashton, do *Massachusetts Institute of Technology* (MIT), em uma apresentação para a Procter & Gamble em 1999. A visão de Ashton para a IoT era de um mundo onde objetos comuns, como embalagens de produtos e utensílios domésticos, poderiam ser equipados com sensores e conectados à internet para fornecer informações em tempo real sobre sua localização, condição e uso (MOUHA, 2021).

Para lidar com esse grande volume de dados, é necessário utilizar técnicas e tecnologias que permitam a coleta, o armazenamento, o processamento e a análise

desses dados de forma integrada, eficiente e escalável. Assim, o termo *Big Data* surgiu para descrever os desafios trazidos pela massiva disponibilidade de dados. O termo *Big Data* e sua definição serão explorados em detalhes mais adiante no artigo. Os dados gerados pelo *Big Data* podem ser utilizados para diferentes fins, como identificar padrões e tendências, realizar previsões, melhorar a tomada de decisão, entre outros. O uso de *Big Data* requer um bom nível de interoperabilidade semântica para que os dados possam ser integrados e utilizados de forma efetiva.

A interoperabilidade semântica é a capacidade de diferentes sistemas compartilharem e entenderem dados e informações entre si, independentemente de como esses dados são armazenados ou processados. Com a interoperabilidade semântica, pode-se integrar dados de diferentes fontes, como sensores, dispositivos móveis, redes sociais e sistemas de gerenciamento de dados, e criar uma visão mais completa e precisa dos dados. Isso é possível graças à utilização de tecnologias semânticas, que permitem que as informações sejam descritas de maneira mais precisa e detalhada, tornando possível sua compreensão e conexão por máquinas e algoritmos.

Assim, faz-se necessário trazer o conceito de *Web* de Dados, que é uma extensão da *World Wide Web* atual que permite que dados sejam compartilhados e reutilizados por aplicativos e usuários, de forma padronizada e estruturada. A ideia é tornar a *Web* mais inteligente, possibilitando que as máquinas entendam e interpretem os dados publicados na Internet, criando uma verdadeira rede de semântica (HEATH; BIZER, 2011; BERNERS-LEE; HENDLER; LASSILA, 2001). A *Web* de Dados, permite que os dados sejam interconectados e compartilhados de forma significativa, tornando-os mais acessíveis e úteis para as diferentes organizações.

De acordo com a W3C (*World Wide Web Consortium*), a *Web* de Dados é composta por três elementos principais: os dados RDF (*Resource Description Framework*), as ontologias representadas em OWL (*Web Ontology Language*) e as consultas SPARQL (*SPARQL Protocol and RDF Query Language*). Esses elementos são usados em conjunto para possibilitar a criação de uma *Web* mais estruturada e interoperável, em que os dados possam ser compartilhados e reutilizados de forma mais eficiente.

Para implementar efetivamente soluções fundamentadas na *Web* de Dados, e fazer um melhor uso dos dados nela disponíveis, é importante ter uma boa gestão de dados. As organizações devem ter um processo de gerenciamento de dados robusto e bem estruturado para garantir que os dados sejam de alta qualidade, íntegros, precisos, relevantes e semanticamente compreensíveis. Isso permitirá que os dados sejam interconectados e compartilhados com outras partes interessadas de maneira eficiente e eficaz (DAMA, 2017; BERNERS-LEE; HENDLER; LASSILA, 2001). A gestão de dados refere-se às práticas e processos usados para gerenciar dados em um ambiente informacional (DAMA, 2017). Isso inclui coletar, armazenar, organizar, proteger e manter os dados, bem como garantir sua qualidade e disponibilidade para uso pelos usuários finais. A organização de dados e informação e a representação da informação são áreas fundamentais da Ciência da Informação (CI).

De acordo com Borko (1968), a Ciência da Informação (CI) tem como objeto de estudo todo o conjunto de conhecimentos relacionados à origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Essa área de conhecimento abrange a pesquisa sobre a representação da informação em sistemas naturais e artificiais, bem como o uso de códigos para a transmissão eficiente da mensagem. Lancaster (1978) ressalta que a organização da informação é essencial para que ela possa ser utilizada como um recurso valioso para a sociedade.

Hjørland (2008) destaca que a organização do conhecimento (OC) está preocupada em desenvolver métodos para organizar e representar o conhecimento de forma que seja acessível e utilizável pelos usuários. Além disso, Hjørland (2008) ressalta a importância da OC para a produção, compartilhamento e disseminação do conhecimento em diferentes áreas do conhecimento, incluindo ciência, tecnologia, medicina e humanidades.

Alvarenga (2003) destaca que a representação do conhecimento é uma questão central na CI, pois está diretamente relacionada à organização e recuperação da informação. Em ambientes digitais, essa questão se torna ainda mais complexa, uma vez que há uma quantidade cada vez maior de informações disponíveis em diferentes formatos e meios. A autora apresenta algumas contribuições da CI para a

representação do conhecimento em ambientes digitais, como o desenvolvimento de padrões e normas para a descrição e organização da informação, a utilização de técnicas de mineração de dados e o estudo de métodos para a representação do conhecimento em diferentes domínios e contextos.

Campos (2004) aborda a importância da modelagem conceitual na representação do conhecimento, enfatizando que essa prática contribui para a organização e a estruturação dos conteúdos informacionais. Ela destaca que o processo de modelização deve levar em consideração os princípios fundamentais da teoria da classificação e da análise de conceitos. Neste trabalho, a autora apresenta uma reflexão teórica e prática sobre a modelagem de domínios de conhecimento na CI, destacando sua importância para a organização e a estruturação dos conteúdos informacionais, bem como para a integração e a recuperação da informação em diferentes contextos.

Considerando os desafios apresentados pelo *Big Data* em lidar com volumes massivos de dados gerados em uma variedade de formatos e por diferentes dispositivos, incluindo a IoT, e considerando ainda a *Web* de Dados como uma fonte importante de dados, torna-se crucial a utilização de mecanismos capazes de interconectar esses dados de forma semântica, a fim de promover a disponibilidade e a capacidade de uso desses dados. Neste contexto, a CI, em especial seus campos de estudo em organização e representação da informação e do conhecimento (ORIC), possui uma relevante contribuição para os desafios trazidos pelo *Big Data*. Ela introduz diversas técnicas de organização e representação da informação que podem ser aplicadas na estruturação e armazenamento de dados e informações, sendo que neste artigo, nosso enfoque será a modelagem de domínio.

A modelagem de domínio (MD) é uma técnica que permite às diferentes organizações a entender e representar seus processos de negócios, sistemas de informação e fluxos de dados. Isso ajuda na identificação de gargalos, otimização de processos e identificação de novas oportunidades de negócios. Ela está intimamente relacionada à ORIC porque seu objetivo é representar, de maneira estruturada e coerente, os conceitos e relações que existem em um determinado domínio de conhecimento ou problema. Os profissionais da informação desempenham um

papel fundamental na modelagem de domínio, pois são responsáveis por entender o domínio, identificar as necessidades do usuário e transformar essas informações em modelos claros e precisos e que sejam semanticamente interoperáveis. Antes de avançar no objetivo do artigo, faz-se necessário delimitar o que se entende como profissional da informação.

Na CI usa-se o termo “profissional da informação” para designar quem tem como objeto de sua atuação, a informação, ou seja, profissionais relacionados ao processamento da informação, tratamento, análise e na disseminação (LOUREIRO; JANNUZZI, 2005). Targino (2000) ao esclarecer sobre o profissional da informação faz uma reflexão sobre algumas premissas necessárias ao entendimento deste conceito. Faz-se necessário compreender o contexto da atuação, a sociedade da informação, os termos informação e profissional, a formação e as TICs. Targino (2000, p. 64) relata que:

o profissional da informação refere-se àqueles que se dedicam à informação, o que implica atualização, capacidade de pesquisa e de manuseio de suportes variados, tendo em vista, sempre, as demandas informacionais do público. É o que justifica, cada vez mais, assegurar que profissional da informação é quem adquire informação registrada, não importa em que tipo de suporte, organiza, descreve, indexa, armazena, recupera e distribui essa informação, tanto em sua forma original, como em produtos elaborados a partir dela, excluindo os produtores de informação, quais sejam, os cientistas e tecnólogos.

Os bibliotecários são historicamente, no Brasil, os profissionais da informação, mas também os são (...) “os documentalistas, arquivistas, museólogos, administradores, contadores, analistas de sistema, comunicólogos, jornalistas, publicitários, estatísticos, engenheiros de sistemas, sociólogos, educadores(...)”, e também as profissões emergentes (TARGINO, 2000).

Este artigo discute as contribuições dos campos da organização e representação da informação com enfoque na atividade de modelagem de domínio com intuito de identificar as principais técnicas, metodologias, práticas, tecnologias e instrumentos de representação necessários para essa atividade. Como contribuição, apresenta-se um conjunto de competências e habilidades necessárias para os profissionais de



informação realizarem as atividades de modelagem de domínios em diferentes contextos informacionais, de forma a ampliarem sua área de atuação para atender as soluções em torno dos desafios trazidos pela Era do *Big Data*. Buscou-se trazer de forma didática todos os conceitos e explicações originários da área da Ciência da Computação, de forma a tornar os aspectos mais técnicos, que não são abordados na formação do profissional da informação.

Conforme a taxonomia da área de CI elaborada por Hawkins, Larson, e Caton (2003) este estudo foca na categoria temática "organização do conhecimento" pois explora questões relacionadas ao tratamento, representação e recuperação da informação. Na tradução adaptada pelas autoras Oddone e Gomes (2003), a categoria temática que este estudo se enquadra é "processamento, recuperação e disseminação da informação".

Em relação aos objetivos, a pesquisa também pode ser classificada como descritiva, exploratória e explicativa. A pesquisa é exploratória, pois visa proporcionar aos pesquisadores uma maior familiaridade sobre o tema pesquisado e os principais conceitos avaliados. Caracteriza-se como Descritiva tem como objetivo descrever as características de determinado fenômeno, no caso os desafios do *Big Data* e as técnicas e práticas de modelagem de domínio, sem buscar estabelecer relações de causa e efeito entre as variáveis estudadas. E ainda é Explicativa pois busca explicar um fato ou fenômeno que é a atuação do profissional da informação na modelagem de domínio (GIL, 2008). Do ponto de vista dos procedimentos técnicos utiliza-se a pesquisa bibliográfica, uma vez que foi elaborada a partir de material já publicado, como livros, artigos e fontes de informação da internet (GIL, 2008).

Na próxima seção busca-se identificar a fundamentação teórica necessária para compreensão do artigo, sendo apresentados conceitos sobre *Big Data* e modelagem de domínio. Na seção 3 são apresentadas algumas das técnicas, metodologias e práticas aplicadas à modelagem. Na seção 4 são sintetizadas as principais competências que o profissional da informação deve desenvolver para atuar com a modelagem de domínio aplicada ao *Big Data*. Na seção 5 são apresentados exemplos de aplicação das práticas do profissional da informação na modelagem de domínio



em ambientes não tradicionais. E por fim, na seção 6 apresentamos uma conclusão com uma breve discussão sobre perspectivas futuras.

## **2 FUNDAMENTAÇÃO TEÓRICA**

A era do *Big Data* trouxe uma grande quantidade de dados disponíveis para as organizações em geral, o que proporcionou a possibilidade de analisar e extrair insights valiosos para a tomada de decisão. Porém, lidar com essa enorme quantidade de informações requer o uso de tecnologias e ferramentas capazes de armazenar, processar e analisar dados. Nesse contexto, a modelagem de domínio se torna uma técnica crucial para a organização e representação dos dados, possibilitando uma melhor compreensão e visualização dos mesmos. Para isso, é importante entender os conceitos de *Big Data*, organização e representação da informação, modelagem conceitual e modelagem de domínio. Esses conceitos são apresentados nesta seção, com o objetivo de fornecer ao leitor uma base para compreender a importância da modelagem de domínio na era do *Big Data*.

Para desenvolver esta seção, utilizou-se artigos e livros indexados na ferramenta de busca Google Acadêmico que indexa, em um único lugar, publicações - artigos, teses, livros, etc. - disponíveis em outras bases. Adicionalmente, foram usadas fontes de referência sobre o tema, identificadas pela expertise das autoras. Na busca foram utilizadas as palavras chaves: *Big Data*, modelagem de domínio, modelagem conceitual, competências informacionais, profissional da informação. Além de pesquisas com combinação destas palavras. A escolha dos documentos, baseou-se: 1º) na ordem de relevância que o Google Acadêmico adota<sup>5</sup> que considera o autor, a publicação do documento, o texto integral de cada artigo, a frequência de citação;

---

<sup>5</sup> Google Acadêmico - “Sobre o Google Acadêmico”. Disponível em: [\[https://scholar.google.com.br/intl/pt-BR/scholar/about.html\]](https://scholar.google.com.br/intl/pt-BR/scholar/about.html). Acesso em: 04/2023.

2º) Na expertise das autoras sobre a literatura da área de CI; 3º) A partir de uma leitura analítica do título e resumo dos artigos selecionados.

## 2.1 Big Data: um termo e muitas definições

O termo *Big Data* já não é uma novidade no meio acadêmico e produtivo, sendo amplamente utilizado para descrever a crescente quantidade de dados gerados e disponíveis para análise. No entanto, apesar da popularidade do termo, ainda não há uma definição universalmente aceita para o conceito de *Big Data*, variando de acordo com o contexto e com o objetivo da análise e da área/campo de estudo inserido. Nesta subseção, são apresentadas algumas das principais definições do conceito de *Big Data*, bem como as suas principais características e desafios associados.

Estamos vivendo em uma era de explosão de dados e informações, impulsionada pelo avanço da tecnologia e da conectividade com o processo de datificação. Diariamente produzimos um número incontável de dados e informações por meio das TICs que nos cercam em nossa vida cotidiana: são dados sobre nossa agenda, nossas preferências, nosso deslocamento, registros sobre educação, saúde, bens e serviços consumidos, lazer, etc. Em paralelo, também demandamos cada vez mais dados para nossas decisões praticamente em tempo real. Somos bombardeados por uma quantidade cada vez maior de informações provenientes de diversas fontes, como redes sociais, dispositivos móveis, sensores e dispositivos conectados à internet das coisas (MAYER-SCHÖNBERGER; CUKIER, 2013).

O grande volume de dados disponíveis hoje em dia, provenientes de diversas fontes e em diferentes formatos, deu origem ao fenômeno conhecido como *Big Data*. O termo foi criado em meados dos anos 90 por Michael Cox e David Ellsworth, cientistas da Nasa, que discutiam os desafios de visualização de grandes volumes de dados, além dos limites computacionais tradicionais para a captura, processamento, análise e armazenamento dessas informações (COX; ELLSWORTH, 1997). O termo foi usado para descrever esse grande conjunto de dados que desafia os métodos e ferramentas tradicionais para manipulação de dados, considerando um tempo razoável de resposta.

De acordo com Taurion (2013), *Big Data* pode ser definido como um conjunto de tecnologias, processos e práticas que possibilitam o armazenamento, gerenciamento e processamento de grandes volumes de dados, sejam eles estruturados ou não, sendo que muitos dos quais anteriormente inacessíveis. Esse tipo de dado, em geral, não pode ser processado pelos sistemas tradicionais de gerenciamento de banco de dados, mas, uma vez processados, as organizações podem utilizá-los para tomadas de decisão mais eficientes ou para melhor gerenciamento de atividades. O autor também destaca que *Big Data* é um termo multidisciplinar que envolve áreas como tecnologia da informação, estatística, matemática, engenharia, ciência da informação, entre outras. Nesta mesma linha, em Amaral (2016), o conceito de *Big Data* é apresentado como uma "nova geração de tecnologias e arquiteturas, projetadas para extrair valor de volumes de dados cada vez maiores, variados e em velocidades crescentes" (p. 17).

Para Laney (2001) existem três características fundamentais associadas ao termo *Big Data*: volume, variedade e velocidade. Tanto Taurion (2013) quanto Amaral (2016) reforçam a presença destas características e apresentam outras: veracidade e valor. O volume diz respeito à quantidade de dados que são produzidos e coletados pelas organizações. As organizações coletam dados de diversas fontes, implicando na variedade dos tipos (estruturados, semiestruturados e não estruturados) e formatos dos dados coletados. A velocidade diz respeito ao quão rápido os dados estão sendo produzidos e quão rápido os dados devem ser tratados para atender a demanda da organização. As decisões são tomadas em tempo real. Temos ainda a veracidade ou confiabilidade dos dados, ou seja, eles devem expressar a realidade e ser consistentes. Enfim, o valor diz respeito à utilidade dos dados ao negócio, ou seja, como agregam valor (TAURION, 2013). Ressalta-se que Sousa *et al* (2018) apontam que a variedade de formatos que os dados são encontrados acarreta em dificuldades de processamento do ponto de vista semântico e demandam modelos de representação que contribuam para a integração destes dados.

Independentemente da definição adotada, o *Big Data* é uma realidade que vem transformando diversos setores, como a indústria, a saúde, a educação, a segurança, entre outros. A análise de grandes volumes de dados pode revelar padrões que não seriam detectados por métodos tradicionais de análise, gerando benefícios para as

organizações e para a sociedade em geral. Com a utilização de *Big Data*, as organizações podem obter insights valiosos a partir de dados antes inexplorados e, assim, tomar decisões mais fundamentadas e eficientes. Entretanto, Souza, Martins e Ramalho (2018) discutem a necessidade de modelos de representação semântica eficientes para lidar com a grande quantidade de dados gerados pelo *Big Data*.

## 2.2 Recuperação da informação

No contexto do *Big Data*, com o grande volume e variedade de dados existentes, é importante recuperar a informação de valor e verdadeira no menor tempo possível. O termo *Information Retrieval* (Recuperação de Informação) foi introduzido por Calvin Mooers em 1951 para determinar uma nova área de estudo. Para Mooers (1951) a recuperação de informação (RI) é entendida como uma área que trata dos aspectos intelectuais da descrição da informação, sua especificação para busca, bem como dos sistemas, técnicas ou máquinas utilizadas para realizar essa operação.

A recuperação de informação é uma área que lida com o armazenamento de documentos e a recuperação automática de informações associadas a esses documentos. RI é o processo de encontrar informações relevantes para uma necessidade de informação a partir de um grande conjunto de dados não estruturados ou semiestruturados. Isso envolve a indexação de documentos, a modelagem de consultas, a recuperação e a apresentação dos resultados relevantes ao usuário. Ela envolve a criação de sistemas de busca eficientes que são capazes de indexar e recuperar informações relevantes a partir de grandes conjuntos de dados, como coleções de documentos ou bancos de dados. A área de RI é multidisciplinar e envolve a combinação de técnicas de ciência da computação, processamento de linguagem natural e teoria da informação para criar sistemas de busca eficientes e precisos (BAEZA-YATES; RIBEIRO-NETO, 2011).

Saracevic (1999) destaca que a RI como a principal área de pesquisa da CI, argumentando que ela é o cerne da atividade de busca e recuperação de informações em ambientes digitais e que a área tem uma ampla gama de questões teóricas e práticas que envolvem a análise e a representação de conteúdos informacionais, bem como a concepção e avaliação de sistemas de RI.

### *2.2.1 Sistemas de recuperação da informação*

Cesarino (1985) apresenta o conceito de sistemas de recuperação da informação como ferramentas que permitem a organização, o armazenamento e a recuperação de informações de forma eficiente. Esses sistemas trabalham com uma grande quantidade de documentos, permitindo que os usuários realizem buscas por meio de palavras-chave ou termos relevantes ao conteúdo desejado.

Sistemas de recuperação da informação (SRI) são sistemas que permitem aos usuários encontrar informações relevantes a partir de uma necessidade de informação expressa por meio de uma consulta. Esses sistemas trabalham com um conjunto de documentos que foram previamente armazenados e indexados, permitindo que os usuários possam recuperá-los de forma eficiente e precisa (BAEZA-YATES; RIBEIRO-NETO, 2011).

Souza (2006) destaca a importância dos SRI no contexto da gestão da informação, ressaltando que esses sistemas têm como principal função a indexação de documentos e a recuperação de informações relevantes para os usuários. Conforme descrito por Souza (2006, p. 163), as atividades de um SRI incluem a disponibilização de informações contidas nos documentos por meio de indexação e descrição, além do armazenamento e gerenciamento físico e/ou lógico desses documentos e suas representações. O sistema deve ainda ser capaz de recuperar as informações e documentos armazenados de maneira a atender às necessidades de informação dos usuários, além de representá-los de forma adequada. Nesta mesma linha, Araújo (2012) aponta que as principais funções de um SRI são representar, armazenar, organizar e localizar itens de informação. A autora também ressalta que a indexação é a principal função de um SRI e que seus componentes incluem documentos, necessidades do usuário, consulta formulada e o processo de recuperação propriamente dito.

Os documentos são considerados objetos informacionais, uma vez que contêm informações que podem ser acessadas e recuperadas por meio de uma busca. Buckland (1991) ao discutir o conceito de informação como coisa, enfatiza a ideia de que os documentos são entidades físicas que contêm informações que podem ser buscadas e recuperadas. Para o autor, a informação deve ser considerada como uma

coisa física, em vez de um conceito abstrato ou intangível. Assim, observa-se os objetos informacionais sendo geralmente compostos de texto, embora possam incluir outros tipos de conteúdo, como imagens, áudios, gráficos e figuras. A organização e representação desses objetos devem permitir que os usuários acessem informações relevantes por meio da expressão de uma necessidade de informação, como uma consulta. Tendo a RI como essência buscar documentos relevantes em resposta a uma determinada consulta que expressa a necessidade de informação do usuário, fornecendo ao usuário uma lista de documentos relevantes, organizados por ordem de relevância, em resposta à sua consulta.

Araújo (2012) versa sobre a comunicação documentária como um processo que ocorre no contexto dos SRI, argumentando que, para que os SRI sejam eficazes, é preciso que haja uma comunicação adequada entre o sistema e o usuário, por meio da mediação dos documentos. Para a autora, a comunicação documentária *é alcançada quando um sistema codifica o conteúdo informacional dos documentos, objetivando sua recuperação pelo usuário* (ARAÚJO, 2012, p. 138).

Com o objetivo de atender às necessidades do usuário aumentando sua satisfação em relação aos SRI, Araújo (2012, p. 139) reflete sobre o papel do bibliotecário, enquanto profissional da informação, como sendo responsável pelo gerenciamento dos SRIs. Para a autora, este profissional deve prestar atenção tanto à forma de representar de maneira precisa e fiel o conteúdo temático dos documentos quanto ao conhecimento das estratégias de busca oferecidas pelo sistema. Os SRIs utilizam linguagens documentárias para ordenar e transmitir a informação, estabelecendo um elo relevante entre os usuários e os próprios sistemas.

### 2.3 Organização e representação do conhecimento

A organização e representação da informação e do conhecimento é um tema central na CI, que busca compreender como as informações podem ser estruturadas, classificadas e descritas para facilitar o acesso e uso pelos usuários. As necessidades dos usuários em acessar as informações se tornam foco central de estudos na área de organização e representação do conhecimento tendo a recuperação da informação como seu motivador.

Bräscher e Café (2008) observaram a falta de clareza na delimitação dos conceitos de "organização da informação" (OI) e "organização do conhecimento" (OC) em diferentes contextos. Em seu estudo, eles propuseram uma definição conceitual na qual a OI é *"um processo que envolve a descrição física e de conteúdo dos objetos informacionais"* (p. 5), resultando na representação da informação como um *"conjunto de elementos descritivos que representam os atributos de um objeto informacional específico"* (p. 5). Por outro lado, a OC tem como objetivo a construção de modelos abstratos da realidade que constituem modelos de mundo.

Buscando definir a área, Hjørland (2017) apresenta a importância da revisão dos conceitos fundamentais da organização do conhecimento. O autor argumenta que a análise crítica desses conceitos é essencial para a compreensão e desenvolvimento da área, destacando os conceitos-chave em organização do conhecimento, como classificação, taxonomia e ontologia, têm implicações significativas na prática da organização do conhecimento. O autor também destaca a importância da relação entre organização do conhecimento e epistemologia, que é o estudo da natureza do conhecimento. Ressalta-se entretanto, que para este artigo, não faz distinção entre organização do conhecimento e organização da informação, pois não é objetivo principal discutir as especificidades entre informação e conhecimento, mas sim apontar questões relacionadas à organização e representação do domínio de informação e conhecimento. Adota-se o termo organização e representação do conhecimento (ORC) para se referir a construção de modelos que representam uma dada realidade.

Segundo Mooers (1951) a organização mecânica do conhecimento tem sido cada vez mais valorizada, especialmente em bibliotecas e centros de documentação. Seu principal objetivo é economizar tempo e recursos humanos, além de melhorar a eficiência do serviço prestado por estas unidades de informação. A ORC visa apoiar o tratamento e recuperação de objetos informacionais de diferentes níveis de estruturação nas organizações. Tem como objetivo tornar as informações acessíveis e úteis para os usuários, permitindo que possam localizar e utilizar o que procuram. Para viabilizar o acesso à informação são necessários aplicar diferentes métodos de organização e representação. O conceito de representar e organizar segundo o



dicionário Cambridge é respectivamente “reproduzir a imagem de algo”<sup>6</sup> e “preparar, planejar”<sup>7</sup>. Pensar em representar é pensar no conceito de substituir ou reproduzir, já a ideia de organizar, nos remete ao fato de fazer algo de forma planejada. Assim, ORC nos leva a ideia de reproduzir ou substituir o conhecimento por conjunto de símbolos mas de maneira sistemática.

A organização e representação da informação e do conhecimento é um aspecto fundamental dos sistemas de recuperação de informação, sendo necessário trazer alguns conceitos para melhor entendimento deste trabalho. Segundo Lima e Alvares (2012, p. 21) *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*. Para estes autores, a representação refere-se às maneiras de expressar a informação e o conhecimento por meio de símbolos ou signos compreensíveis para usuários e sistemas de informação.

De acordo com Taylor (2004, p. 1), a organização é essencial para a recuperação da informação, especialmente quando há uma grande quantidade de objetos informacionais para serem consultados. A representação da informação é utilizada para facilitar a identificação e acesso a esses objetos, e recursos de organização e recuperação da informação são empregados com o objetivo de conectar as necessidades de informação dos usuários com as informações relevantes armazenadas nos sistemas de informação.

Para Alvarenga (2003, 2006) a representação do conhecimento é o processo pelo qual ocorre a transformação do conhecimento em uma forma que possa ser comunicada, armazenada e processada, e envolve a escolha de uma linguagem ou sistema de símbolos que permita a expressão do conhecimento de forma clara e precisa.

A representação da informação se refere à maneira como os itens em uma coleção são descritos e organizados para que possam ser recuperados de maneira eficaz. A

---

<sup>6</sup> Definição disponível em: <https://dictionary.cambridge.org/pt/dicionario/portugues-ingles/representar>

<sup>7</sup> Definição disponível em: <https://dictionary.cambridge.org/pt/dicionario/portugues-ingles/organizar>

maioria dos SRI representa os itens em sua coleção usando palavras. Em linhas gerais, as palavras fornecidas pelo usuário são então comparadas às palavras anexadas aos itens na coleção. Se as palavras do usuário corresponderem às do item, esse item pode ser o que o usuário deseja. O processo de organização pode ser encarado como um tipo de uso de uma linguagem específica, permitindo assim utilizar construtos linguísticos como vocabulário, semântica e sintaxe para ampliar a compreensão e avaliar diferentes métodos de organização. Essa abordagem também possibilita uma conceitualização mais ampla, que pode unificar métodos antes considerados distintos, como a catalogação, classificação e indexação (SVENONIUS, 2000, p. 6). De acordo com Svenonius (2000), para que a informação possa ser organizada, é necessário que ela seja descrita. Ele explica que uma descrição é uma afirmação sobre as propriedades de um objeto ou sobre suas relações com outros objetos, o que permite identificá-lo.

A OCR se dedica a encontrar teorias e ferramentas que aprimorem as formas de armazenamento e recuperação de informações, que por sua vez exigem representação. Assim, a ideia de organização de conteúdos informacionais envolve a criação de métodos e instrumentos para representar informações por meio de conceitos, com destaque para as linguagens documentárias e suas variações, como sistemas de classificação, taxonomias e ontologias (CINTRA *et al.*, 2002). O objetivo da OCR é alcançado por meio da descrição física e de conteúdo dos objetos informacionais. A descrição de conteúdo refere-se ao conhecimento registrado, enquanto a descrição física se relaciona ao suporte da informação. O resultado dessa descrição é a representação da informação, que consiste em um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico.

Hjørland (2007) discute as relações entre semântica e organização do conhecimento e defende que a semântica é essencial para a organização do conhecimento e que uma teoria de semântica deve estar no centro das preocupações da disciplina. Para o autor, a semântica deve ser entendida como o estudo da relação entre sinais (palavras, símbolos) e os objetos ou conceitos aos quais se referem.

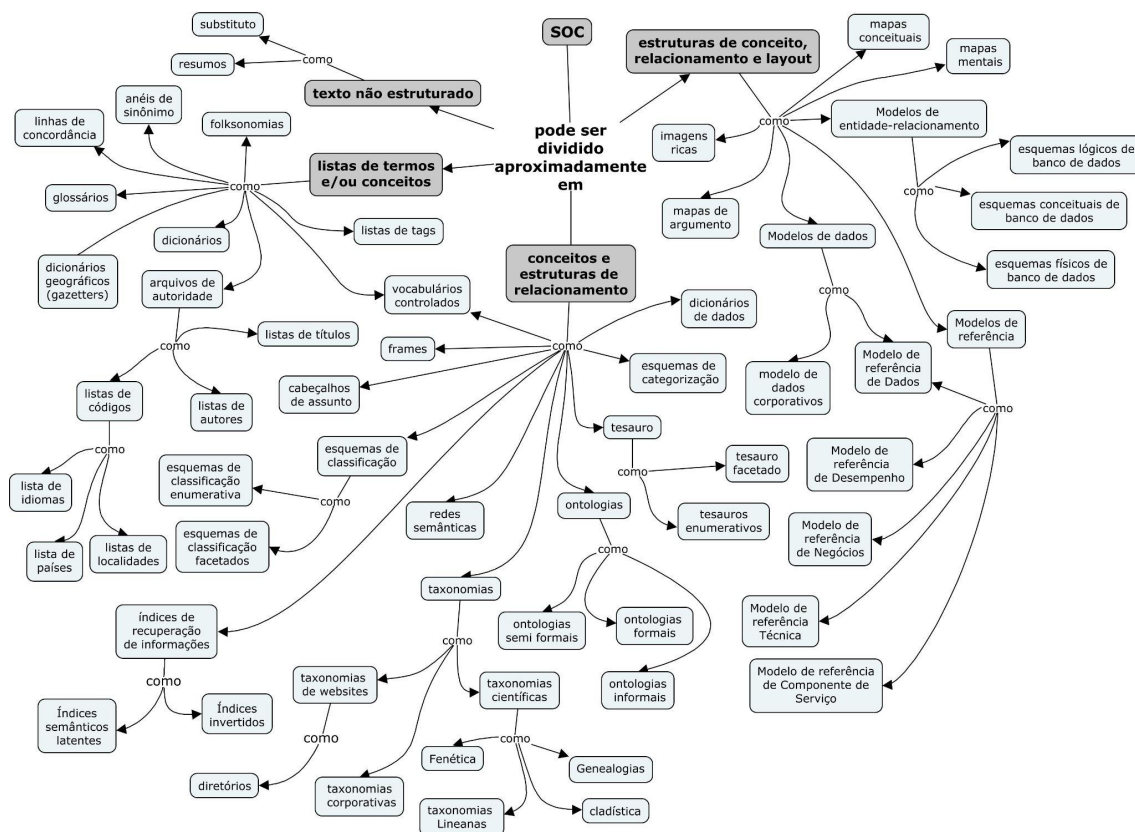
### *2.3.1 Sistema de Organização do Conhecimento*

Sistema de Organização do Conhecimento (SOC) são sistemas compostos por conjuntos de termos, conceitos e outros elementos de informação que são usados para classificar, categorizar, indexar e representar diferentes tipos de conhecimento, tornando mais fácil a recuperação e a navegação através de recursos de informação, como bases de dados, bibliotecas digitais e outros ambientes informacionais (MAZZOCCHI, 2018). Uma contribuição a partir de Taylor (2004) é a definição de SOC como um conjunto organizado de categorias, classes ou termos utilizados para representar e agrupar o conhecimento de uma área ou domínio específico.

Em síntese, os SOCs são sistemas conceituais que ajudam a organizar o conhecimento para que possa ser melhor utilizado e compartilhado entre diferentes usuários e sistemas por meio da representação de um determinado domínio onde ocorre uma sistematização dos conceitos e das relações semânticas que se estabelecem entre eles.

Souza, Tudhope e Almeida (2010) apresentam uma tipologia de SOC (Figura 1) baseada em um modelo de quatro dimensões: controle de vocabulário, estruturação do conhecimento, nível de agregação e especialização. Essa tipologia pode ser usada como uma ferramenta para compreender as diferenças e semelhanças entre diferentes tipos de SOC, permitindo uma melhor seleção e utilização desses sistemas de acordo com as necessidades específicas de cada contexto de informação.

Figura 1: Tipologia de Sistema de Organização do Conhecimento



Fonte: Traduzido da Fig. 1 de Souza, Tudhope e Almeida (2010)

Nota-se que conforme esta tipologia, SOC's englobam modelos de dados, modelo de entidade-relacionamento, glossários, folksonomias, sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, tesaurus, taxonomias e ontologias, de fato, todos estes artefatos organizam e representam o conhecimento. Nota-se ainda que alguns tipos são mais adequados para representar conceitos e as relações semânticas entre estes conceitos.

### 2.4 Modelagem de domínio: conceitos e instrumentos de representação

Os modelos, de maneira geral, visam formalizar o universo por meio de expressões controláveis pelo ser humano, sendo uma resposta à necessidade humana de compreender a aparente complexidade do mundo que nos rodeia (SAYÃO, 2001).

A modelagem conceitual teve sua origem na década de 1970, quando Peter P. Chen desenvolveu estudos sobre a abordagem de entidade-relacionamento na modelagem de dados. Essa abordagem proposta por Chen consiste em representar as entidades (objetos do mundo real) e seus relacionamentos em um modelo

conceitual. Ao longo dos anos, a modelagem conceitual evoluiu e incorporou novos elementos, tornando-a mais rica em termos de semântica e aplicabilidade em diferentes contextos. Essa evolução permitiu que a modelagem conceitual fosse utilizada não apenas para representar a estrutura de dados de um sistema, mas também para capturar aspectos comportamentais, restrições e regras de negócio (COUGO, 1997).

A representação da informação e a modelagem de domínio são duas áreas interconectadas na CI. A representação da informação refere-se às técnicas e processos usados para organizar e representar informações de forma significativa e acessível. Já a modelagem de domínio envolve a criação de modelos conceituais que descrevem os elementos e relacionamentos em um domínio específico. Para Bräscher e Café (2008) o processo de organização do conhecimento se baseia na análise dos conceitos e de suas características, com o objetivo de estabelecer a posição de cada conceito dentro de um determinado domínio, bem como de suas relações com os demais conceitos que compõem esse sistema.

Para entender a ideia por trás do conceito, resgata-se o trabalho de Dahlberg (1978) que explora as bases filosóficas do conceito e sua relação com a linguagem, o conhecimento e a comunicação, apresenta uma análise detalhada das várias dimensões do conceito e propõe uma estrutura abrangente para seu estudo. De acordo com Dahlberg (1978), o conceito é um elemento fundamental da comunicação e do conhecimento, sendo considerado uma construção social que emerge da interação entre a linguagem, a mente e o mundo. Complementando, Abbagnano (2003, p. 164, citado por Kobashi e Francelin, 2011) descreve conceito como um processo que possibilita a descrição, classificação e previsão de objetos cognoscíveis, incluindo qualquer tipo de sinal ou procedimento semântico, independentemente do objeto a que se refere. Além disso, os conceitos podem ser classificados quanto à sua natureza, que pode estar na essência ou no signo, e quanto à sua função, que pode ser final ou instrumental.

Para explorar o termo “modelo de domínio”, encontramos o trabalho de Campos (2004) que aborda os princípios fundamentais para a modelagem de domínio, com base em uma investigação que envolveu a análise de trabalhos científicos e

entrevistas com especialistas da área. A autora destaca a importância da definição clara dos objetivos e escopo do domínio em questão, bem como da identificação dos conceitos, relações e regras que o compõem. Campos e Gomes (2014) que discutem os princípios fundamentais para a modelagem de domínio a partir da visão de dois importantes autores da área: Barry Smith e Ingetraut Dahlberg. O trabalho destaca a importância da identificação e representação dos conceitos, relações e regras que compõem o domínio, bem como da definição de ontologias e taxonomias que permitam uma organização eficiente do conhecimento. Tanto Campos (2004) quanto Campos e Gomes (2014) enfatizam a necessidade de uma abordagem colaborativa e inter-multidisciplinar na modelagem de domínio, que envolva profissionais de diferentes áreas do conhecimento, tais como informática, ciência da informação e epistemologia, entre outras.

A partir de um levantamento conceitual de diversos trabalhos - Campos (2004); Campos e Gomes (2014); Chen (1990); Cougo (1997) - conclui-se que a modelagem de domínios refere-se ao processo de construir abstrações de alto nível sobre um domínio de conhecimento a fim de possibilitar a reutilização dos conceitos e relações entre os conceitos representados no artefato construído. Este artefato busca representar de forma simplificada e abstrata os principais aspectos do mundo real relacionados a um determinado domínio de aplicação.

Resgata-se ainda Sayão (2001) que descreve os modelos conceituais como representações simplificadas e compreensíveis do mundo, que nos possibilita visualizar as características essenciais de um determinado domínio ou campo de estudo. Esses modelos são aproximações da realidade em que detalhes acidentais são eliminados, permitindo que aspectos essenciais e relevantes do mundo real se tornem visíveis, com o objetivo de transmitir informações sobre ele. O autor aborda a importância dos modelos teóricos na CI. O autor discute o processo de abstração e a aplicação do método científico na construção de modelos teóricos nessa área.

Cabe esclarecer que embora os termos "modelo de domínio" e "modelo conceitual" sejam frequentemente usados de forma intercambiável, eles podem ter significados ligeiramente diferentes, dependendo do contexto. Em geral, um modelo de domínio refere-se a uma representação abstrata e simplificada de um determinado campo de

conhecimento ou área de interesse. Ele descreve as entidades, os conceitos, os relacionamentos e as regras de negócio relevantes nesse domínio específico e são usados para capturar o conhecimento essencial do domínio e fornecer uma estrutura para entender e comunicar as características e os requisitos desse domínio. Por outro lado, um modelo conceitual também representa uma visão abstrata de um sistema ou de um conjunto de conceitos, e é usado para capturar a estrutura, os componentes e os relacionamentos entre os conceitos em um determinado contexto. Um modelo conceitual pode ser aplicado a vários domínios, sendo mais abrangente e geral em sua representação. Ele é frequentemente usado para ilustrar os conceitos-chave e as interações entre eles. Este artigo adotará os termos "modelo de domínio" e "modelo conceitual" como sinônimos. Embora eles possam ter algumas distinções sutis em diferentes contextos, neste texto, serão usados de forma intercambiável para se referir a representações abstratas e simplificadas de um campo de conhecimento específico. Ambos os termos descreverão a estrutura, as entidades, os conceitos e os relacionamentos relevantes em um determinado domínio ou contexto.

A modelagem de domínio ou conceitual possui vantagens em relação à linguagem natural devido à sua capacidade de oferecer um reconhecimento preciso das semânticas da aplicação por se basear em uma notação formal. Enquanto a linguagem natural pode ser ambígua e sujeita a interpretações variadas, a modelagem conceitual utiliza uma linguagem visual ou simbólica para representar os conceitos e suas relações de forma clara e precisa. Os seres humanos têm a habilidade natural de elaborar modelos como uma forma de compreender o mundo ao seu redor, estabelecer padrões de comunicação e representar de maneira simplificada objetos ou situações complexas. Os modelos conceituais são construídos a partir de abstrações que destacam os relacionamentos entre conceitos, explorando similaridades, diferenças e outras associações de significado. Ao utilizar a modelagem conceitual, é possível representar visualmente as relações entre os conceitos de um domínio específico, tornando mais fácil identificar padrões, interações e estruturas subjacentes. Essa representação visual facilita a comunicação e a compreensão entre os envolvidos, uma vez que os modelos



conceituais fornecem uma linguagem comum e uma representação visual que transcende as barreiras da linguagem natural (SALES; SAYÃO; MOTTA, 2012).

#### *2.4.1 Modelos e Modelagem*

Os modelos e a modelagem são elementos fundamentais em diversas áreas de estudo e prática, incluindo a CI, a engenharia, a computação, entre outras. Os modelos são representações simplificadas e abstratas da realidade, que permitem compreender e estudar fenômenos complexos de forma mais acessível e controlada. Eles podem ser utilizados para diferentes propósitos, como comunicar informações, prever resultados, simular situações e auxiliar na tomada de decisões.

A modelagem, por sua vez, é o processo de criar e desenvolver modelos. Envolve a seleção de elementos relevantes do objeto ou sistema a ser modelado, a definição de suas relações e interações, e a formulação de regras e princípios que regem o comportamento do modelo. A modelagem pode ser realizada de forma manual, utilizando técnicas gráficas ou matemáticas, ou de maneira computacional, com o auxílio de software e algoritmos específicos.

Um modelo tem diversos propósitos, mas sua função primordial é comunicar informações sobre o objeto da modelagem, a fim de proporcionar um entendimento mais abrangente da realidade. A atividade de modelagem, por sua vez, requer que aquele que está realizando a modelagem possua uma visão clara e livre de ambiguidades do sujeito ou objeto sendo modelado. Além disso, é necessário fazer uma seleção adequada dos elementos do universo do discurso que comporão a visão a ser representada (SAYÃO, 2001).

Entre outros trabalhos, Sayão (2011) cita o trabalho de Herbert Stachowiak que apresenta três características básicas dos modelos:

**a) Característica de mapeamento:** Os modelos são representações de "originais" ou "protótipos", que podem ser naturais ou artificiais. Eles têm a função de modelar algo existente no mundo real.

**b) Característica de redução:** Os modelos geralmente não representam todos os atributos do original que eles representam, mas apenas aqueles que são relevantes para quem está construindo o modelo. Isso implica em uma simplificação e seleção dos elementos essenciais do objeto modelado.

**c) Característica de pragmatismo:** Os modelos não são idênticos aos seus originais. Eles são construídos com a finalidade de substituição e desempenham funções específicas, orientadas para objetivos mentais ou factuais, em um determinado contexto e período de tempo limitados. Os modelos são utilizados para operações práticas e têm uma finalidade utilitária.

A criação de modelos envolve a manipulação de características, como agrupamento, descarte ou identificação de características comuns entre os objetos. Esses modelos desempenham um papel importante na construção de nossa concepção e compreensão da realidade, bem como na expressão, comunicação e aprendizado de informações sobre essa realidade.

Modelar domínios envolve a organização e representação de informação, provenientes de diversas fontes independente do formato, já que a visão deste modelo não enfatiza as características da solução tecnológica mas sim a visão semântica da informação. O objetivo é ter uma semântica padronizada e interoperável, possibilitando a troca de informação entre os diferentes agentes que conseguem trabalhar com estas visões conceituais. A seguir citamos alguns SOCs que visam a modelagem conceitual, conforme a taxonomia proposta na Figura 1.

#### *2.4.2 Sistemas de Vocabulários Controlados, Taxonomias, Tesouros e Ontologias*

Entre os sistemas de organização do conhecimento utilizados na modelagem de domínios, listamos a seguir alguns artefatos informacionais, seus conceitos e suas principais características:

**a) Vocabulário controlado:** é uma lista de termos autorizados que apresenta uma estrutura semântica. Esses termos são utilizados pelo indexador na análise de assunto para classificar um documento. Tem como objetivo: 1) Controlar o uso de sinônimos. 2) Diferenciar termos com mesma grafia e significados diferentes. 3) Agrupar termos por meio de relacionamentos hierárquicos (LANCASTER, 2014). Para fins de modelagem realiza o controle de sinônimos entre os termos;

**b) Taxonomia:** é um termo originado do sistema de classificação hierárquica na área de zoologia e botânica no século XVIII, desenvolvido por Carl Linnæus. Na CI é definido como uma classificação hierárquica de assuntos com função de organização lógica. A taxonomia não apresenta a definição dos termos e sim as relações hierárquicas entre eles. Para fins de modelagem apresenta entrada única e hierarquizada dos termos (SOUZA, 2015);

**c) Tesouro:** Para Currás (1995, p. 88) o tesouro é “uma linguagem especializada, normalizada pós-coordenada, usada com fins documentários, onde os elementos lingüísticos que o compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente”. Podem ser monolíngüe, bilíngues ou plurilíngües fornece uma estrutura rica para explicar os termos, descreve os assuntos de forma hierárquica e apresenta notações para designar os termos, como: BD: *broader term* - termo mais amplo; SN: *scope note* - nota de escopo; USE – para indicar o termo sinônimo permitido; TT: *top term* - termo principal; RT: *related term* – termo relacionado. Tem como objetivo a recuperação da informação (SOUZA, 2015). Para fins de modelagem apresenta os termos em hierarquia e notação sobre os termos;

**d) Ontologia:** Segundo Almeida (2020, p. 23), “é um termo que refere a um artefato representacional, o qual inserido em sistemas, provê uma estrutura de mundo à máquinas permitindo-lhes um tipo básico de raciocínio”. A ontologia apresenta a descrição formal dos conceitos e as relações entre eles por meios de axiomas, com a finalidade de permitir o compartilhamento e reutilização de conhecimento. Tem como objetivo descrever um domínio do conhecimento, ou seja, modelar a realidade. Assim este artefato informacional para fins de modelagem apresenta definições precisas e formais de relacionamentos entre classes (conceitos) e instâncias dessas classes (SOUZA, 2015).

#### *2.4.3 Modelos conceituais de dados*

Segundo Campos (2004), um modelo conceitual é uma representação que facilita a comunicação, permitindo que se desloque a concepção de uma análise subjetiva sobre o real. Ele possibilita a compreensão das características do objeto de estudo, proporcionando uma visão mais clara e estruturada.

Uma importante contribuição para a ideia de modelos de domínio aplicados ao *Big Data* remete à modelagem de dados ou modelagem conceitual de dados originária na Ciência da Computação (CC). Os modelos de dados são um conjunto de conceitos que são organizados para descrever como um banco de dados vai ser criado.

A visão sobre modelos conceituais remetem a Chen (1990) que apresenta uma introdução completa à modelagem de dados com foco na abordagem entidade-relacionamento. Outra contribuição da área de banco de dados é o trabalho de Cougo (1997) que apresenta os principais conceitos relacionados à modelagem conceitual e ao projeto de banco de dados, explorando desde as etapas iniciais de análise de requisitos até a implementação física do banco de dados.

A identificação de entidades e relacionamentos na abordagem entidade-relacionamento é considerada uma etapa crucial na modelagem de dados, pois permite a captura da semântica dos dados. Através dessa identificação, é possível entender as entidades envolvidas em um determinado domínio, bem como as relações e conexões entre elas. Com isso, é possível construir modelos conceituais que refletem de forma mais precisa a realidade do domínio em questão, facilitando o processo de desenvolvimento de sistemas e bancos de dados (CHEN, 1990; COUGO, 1997).

Entretanto, a abordagem ER pode ser incrementada pelas teorias da CI. Maia e Alvarenga (2014) apresentam uma análise sobre a aplicação da teoria da classificação facetada no modelo ER. As autoras destacam que a classificação facetada oferece uma metodologia sistemática e rigorosa para a organização de informações em bibliotecas, museus e outras instituições, e que esses princípios podem ser aplicados na modelagem de bancos de dados. A partir dessa análise, as autoras propõem algumas contribuições para o modelo ER, como a utilização de facetas na representação de entidades e atributos, a incorporação de relações hierárquicas entre as entidades, e a utilização de regras de negócio baseadas na classificação facetada para a validação de dados.

A Teoria da Classificação Facetada é um modelo teórico de organização do conhecimento, proposto pelo bibliotecário indiano Ranganathan. Ela propõe que a organização do conhecimento deve ser feita por meio da identificação de características fundamentais dos objetos, chamadas de facetas, que devem ser combinadas para criar uma classificação completa e coerente (DAHLBERG, 1972).

### **3 PRÁTICAS PARA MODELAGEM DE DOMÍNIOS**

A modelagem de domínios é um processo fundamental para a organização do conhecimento e a representação de informações de maneira clara e precisa. Existem diversas práticas e metodologias disponíveis para essa tarefa, e é importante conhecer suas características e aplicabilidades para escolher a mais adequada para cada contexto. Nesta seção, exploraremos algumas das principais práticas para modelagem de domínios, divididas em subseções que abordam desde as definições aristotélicas e hierarquias, passando por metodologias para construção de

ontologias e modelos Entidade-Relacionamento, até as linguagens de marcação e tecnologias e ferramentas para representação. Cada uma dessas subseções trará uma visão geral e exemplos práticos das respectivas práticas, fornecendo um panorama completo das opções disponíveis para modelagem de domínios.

### 3.1 Modelo Entidade-Relacionamento

A abordagem entidade-relacionamento (ER) utiliza diagramas para representar as entidades e seus relacionamentos, bem como os atributos das entidades. Uma *entidade* é vista como um objeto ou coisa no mundo real, como pessoas e lugares, que pode ser identificado de forma única em relação a outros objetos. Essas entidades podem ser concretas ou abstratas e possuem *atributos* que são características ou propriedades descritivas de cada membro do conjunto de entidades. Além disso, uma entidade pode se associar com outra entidade, assim são considerados os *relacionamentos*, que são estas associações entre uma ou várias entidades. Os relacionamentos descrevem como as entidades se relacionam entre si, podendo ser um para um, um para muitos ou muitos para muitos (CHEN, 1990, p. 21-24).

Cougo (1997) aborda a modelagem conceitual de dados, apresentando conceitos como entidades, relacionamentos e atributos, além de técnicas como o diagrama Entidade-Relacionamento (DER), e aborda questões fundamentais como a identificação de entidades e atributos, relacionamentos e cardinalidades, bem como técnicas avançadas de modelagem de dados, como modelagem de herança, agregação e generalização.

### 3.2 Ontologias como Modelos Conceituais

As ontologias são sistemas de organização do conhecimento e desempenham um papel importante como uma abordagem específica de modelagem conceitual. Uma ontologia pode ser definida como uma representação formal e explícita de conceitos, suas propriedades e as relações entre eles em um determinado domínio de conhecimento. Ela busca capturar a essência e a estrutura conceitual desse domínio, fornecendo uma base sólida para a organização, compartilhamento e compreensão de informações. Uma das vantagens das ontologias como prática de modelagem conceitual é a capacidade de fornecer uma estrutura clara e bem definida para o

domínio em questão. Elas permitem a criação de um vocabulário comum e compartilhado, facilitando a comunicação e o entendimento mútuo entre os diferentes atores envolvidos. Além disso, as ontologias permitem a inferência automática de novas informações com base nas relações estabelecidas, ampliando assim o poder de representação e o potencial de descoberta de conhecimento (ALMEIDA, 2020; 2021a; FARINELLI; SOUZA, 2012; FARINELLI;ALMEIDA, 2019; SILVA et al., 2014; FARINELLI et al., 2013).

A construção de uma ontologia envolve a identificação dos conceitos relevantes do domínio, suas propriedades e as relações entre eles. Essa tarefa requer uma análise profunda do domínio de estudo, envolvendo a participação de especialistas e a revisão de literatura. A ontologia resultante é uma representação formal e semântica do conhecimento, que pode ser utilizada para diversos fins, como a organização de bases de dados, a interoperabilidade entre sistemas e a busca de informações (FARINELLI; ALMEIDA, 2019).

No campo da Ciência da Informação, as ontologias têm sido amplamente aplicadas em áreas como a recuperação de informação, a organização de conteúdo, a gestão do conhecimento e a construção de sistemas inteligentes. Elas fornecem uma estrutura sólida para a representação e o acesso a informações, permitindo uma melhor compreensão dos dados e uma maior precisão na busca e no processamento de informações (ALMEIDA, 2021b; FARINELLI et al., 2013).

### *3.3.1 Metodologias para construção de ontologias*

Na construção de artefatos ontológicos, Schulz et al. (2012) apresentam diretrizes que orientam a construção de bons artefatos:

- Os artefatos ontológicos devem ser formais, ou seja, suas entidades (termos) são representadas sem ambiguidade de significado e a especificação das relações entre os termos incluem axiomas lógicos, permitindo a automatização tanto da recuperação quanto do raciocínio.
- As ontologias devem apresentar especificações explícitas permitindo que as pessoas raciocinem e entendam o domínio representado pela ontologia.
- As ontologias devem ser adequadas ao domínio alvo da representação, significa que a ontologia construída deve representar a realidade e o conhecimento presentes no domínio desejado.

Adicionalmente, a construção de bons artefatos ontológicos deve seguir algumas metodologias para o desenvolvimento, não existe um consenso sobre qual metodologia é mais apropriada. Farinelli (2017) apresenta uma revisão sistemática da literatura, buscando uma metodologia adequada para construir a ontologia OntONEo. Como parte desta revisão, foram analisadas as metodologias mais conhecidas de meados dos anos 90 até 2016, algumas dessas metodologias são: Toronto Virtual Enterprise (TOVE); Methontology; Metodologia NeOn (Network Ontology); Systematic Approach for Building Ontologies (SABiO); Up for ONtology; Metodologia do Realismo Ontológico.

#### **a) Metodologia Ontoforinfoscience**

Mendonça (2015) descreve uma nova metodologia de construção de ontologias no âmbito da CI, chamada *OntoForInfoScience*, que se diferencia por detalhar as etapas e atividades do ciclo de desenvolvimento ontológico e que enfatiza a importância dos aspectos filosóficos e teóricos envolvidos no processo.

O objetivo da *OntoForInfoScience* é auxiliar especialistas em ORC, incluindo cientistas da informação, a lidar com questões lógicas e filosóficas que envolvem o desenvolvimento de ontologias. Embora a solução desses problemas exija outras ações, como a inclusão de requisitos de formação técnica-acadêmica, a metodologia *OntoForInfoScience* representa um passo em direção a um melhor entendimento dos termos e detalhes técnicos do processo de desenvolvimento de ontologias por parte dos cientistas da informação. A metodologia foi criada durante o desenvolvimento da ontologia HEMONTO, que faz parte do projeto Blood, que busca gerar uma linguagem formal sobre o sangue humano no domínio da hematologia e hemoterapia (MENDONÇA, 2015; MENDONÇA; ALMEIDA, 2015; MENDONÇA; SOARES, 2017).

#### **b) 3.3.2 Metodologia ReBORM**

A metodologia ReBORM (*Realism-Based Ontology Engineering Methodology*) é uma abordagem de engenharia de ontologia usada para a construção de uma ontologia aplicada ao domínio obstétrico. A metodologia ReBORM combina aspectos da metodologia NeOn e da metodologia do realismo ontológico para a construção de ontologias de forma coerente e bem documentada. A NeOn fornece um conjunto de



melhores práticas para a engenharia de ontologia, incluindo nove cenários que destacam a reutilização de recursos ontológicos e não ontológicos, bem como um arcabouço de 59 processos e atividades para a construção de artefatos ontológicos. ReBORM adota princípios e aspectos da metodologia do realismo ontológico para garantir que os artefatos ontológicos representem o conhecimento sem ambiguidades, maximizando a coerência semântica para humanos e computadores. Do ponto de vista prático, ReBORM, combina as etapas da metodologia do realismo ontológico com os cenários e atividades da metodologia NeOn. O desenvolvimento da uma ontologia de domínio deve seguir um ciclo de vida de desenvolvimento, iterativo-incremental (FARINELLI, 2020).

Para conhecer detalhes desta metodologia, sugere-se a leitura de Farinelli (2020), Farinelli (2017) e Farinelli e Elkin (2017). Estes trabalhos explicam as etapas do processo de construção da ontologia, que inclui atividades como definição do escopo, a identificação e a seleção dos termos relevantes do domínio, a criação das entidades/classes, definição de propriedades e axiomas, e a validação da ontologia.

### 3.4. Hierarquias e definições aristotélicas

Na modelagem de domínios um requisito importante é a definição dos termos. A formulação de definição de termos desempenha uma importante aplicação em sistemas de terminologias, classificações, tesouros e ontologias. As definições servem para uma variedade de funções como: 1) Aumentar e enriquecer vocabulário; 2) Introduzir o significado e uso de novas palavras; 3) Eliminar ambiguidade; 4) Reduzir a imprecisão; 5) Solucionar problemas epistemológicos. (SOUZA, 2015).

Para a elaboração de definições na modelagem de domínios é preciso seguir princípios Aristotélicos. A definição aristotélica é composta pela combinação de *Genus* (gênero, primeira parte da definição) + *Differentia* (differential/diferença, segunda parte). O *Genus* significa qual é a característica essencial para os subtermos de uma hierarquia e a *Differentia* são as características essenciais que diferem um termo de outro termo. (SOUZA; ALMEIDA, 2022; SOUZA; FARINELLI; FELIPE, 2021).

Um princípio importante nas definições aristotélicas é o de herança única. A definição de uma classe na ontologia precisa especificar quais as características desta classe de forma consistente para que seja assegurada a sua herança transitiva pela hierarquia, ou seja, a classe superior passar características para as subclasses. A definição de um conceito na ontologia agrega características das definições dos termos superiores da hierarquia, ou seja, as características essenciais de um termo passam para os subtermos. Assim, no princípio de herança única: o *Genus* e a *Differentia*, são importantes. O *Genus* ou herança é a característica essencial que todos os subtermos da hierarquia irão assumir e *Differentia* é a característica essencial que distingue um termo do outro. (SOUZA, 2015; SOUZA; ALMEIDA, 2022).

A consistência nas definições necessita de um ponto de vista unificador (ou seja, o contexto) na representação do conceito. Esse contexto deve manter-se em toda a ontologia. Ao obedecer a esse requisito, a posição de um conceito na hierarquia será enriquecida pela sua própria definição e pela definição do termo superior dentro dessa hierarquia. (SOUZA, 2015; SOUZA; ALMEIDA, 2022). Exemplos práticos de como formular definições aristotélicas são demonstrados no método OntoDef (SOUZA; ALMEIDA, 2022; SOUZA; FARINELLI; FELIPE, 2021).

### 3.5 Linguagens de marcação

Uma linguagem de marcação é um conjunto de convenções que especifica a codificação de texto, onde as marcações são utilizadas para representar informações estruturadas (ALMEIDA, 2002). Linguagem de marcação é uma forma de codificação de informações que utiliza tags (etiquetas) para identificar elementos em um documento de texto, indicando como esses elementos devem ser exibidos ou processados. Segundo Gilliland (2002), as linguagens de marcação permitem a separação clara do conteúdo do documento da sua formatação, o que facilita o armazenamento, a transmissão e o processamento das informações em diferentes contextos e plataformas.

As linguagens de marcação são amplamente utilizadas na CI, principalmente para a criação e organização de recursos eletrônicos, como páginas da *Web*, documentos XML e metadados. Ma e Xiao (2011) destacam que as linguagens de marcação mais comuns incluem HTML (*HyperText Markup Language*), XML (*Extensible Markup*

*Language*) e RDF (*Resource Description Framework*). Adicionalmente, temos a OWL (*Web Ontology Language*) para a descrição e definição de ontologias. O HTML e o XML são linguagens mais amplamente utilizadas na *Web* para a marcação de conteúdo, enquanto o OWL e o RDF são utilizados em sistemas de organização de conhecimento e *Web Semântica* (GUIMARÃES et al., 2018).

Sousa et al. (2018) apresentam o contexto de uso destas tecnologias - RDF e OWL, além do SKOS<sup>8</sup> - e como elas são utilizadas para representar as informações no ambiente *Web* no contexto do *Big Data*.

### **a) *HyperText Markup Language (HTML)***

HTML é a linguagem de marcação padrão utilizada para criar páginas *Web* e estruturar seu conteúdo. É responsável pela estruturação e organização de documentos hipertexto, permitindo que elementos multimídia como imagens, áudio e vídeo sejam integrados a uma página *Web*. Além disso, ele é a base para a criação de sites responsivos, ou seja, que se adaptam a diferentes tamanhos de tela, o que é muito importante em um contexto onde o acesso à internet é feito cada vez mais por dispositivos móveis (COREMANS, 2015).

De acordo com Rezende (2011), as bibliotecas podem utilizar o HTML para criar seus próprios sites e portais de busca, facilitando o acesso à informação por parte de seus usuários. Além disso, o HTML pode ser utilizado para criar páginas de descrição de coleções e acervos, para disponibilizar catálogos online e para criar interfaces de busca mais amigáveis. No contexto do trabalho do bibliotecário, o HTML é uma ferramenta essencial para a criação e organização de conteúdo na *Web*.

O HTML é a linguagem de marcação mais utilizada para a criação de páginas da *Web*, enquanto o XML é usado principalmente para a representação de dados estruturados e documentos de intercâmbio de informações.

---

<sup>8</sup> Simple Knowledge Organization System

### **b) Extensible Markup Language (XML)**

A linguagem de marcação Extensible Markup Language (XML), assim como sua precursora SGML “é um padrão internacional, não proprietário e de código aberto, utilizado já há bastante tempo para troca eletrônica de dados e que pode ser utilizada por diferentes sistemas informatizados” (ALMEIDA, 2002, p. 6).

XML é uma linguagem de marcação que permite a codificação de documentos com a finalidade de armazenamento e transmissão de dados. Ela define um conjunto de regras para a criação de tags (etiquetas) que são utilizadas para identificar e descrever o conteúdo de um documento. Uma das vantagens do XML é a possibilidade de o autor do documento definir suas próprias tags, conferindo ao conteúdo habilidades semânticas que permitem melhorias significativas em processos de recuperação e disseminação da informação (BRAY et al., 2008).

Essa linguagem foi criada no final do anos 90 pelo *World Wide Web Consortium* (W3C) com o objetivo de permitir a troca de informações entre diferentes sistemas de forma mais flexível e interoperável. O XML é uma linguagem de marcação que tem sido amplamente utilizada em sistemas de informação e gerenciamento de conteúdo em diferentes áreas, incluindo a Biblioteconomia. O uso do XML em bibliotecas e centros de informação, por exemplo, tem se tornado cada vez mais comum, pois a linguagem permite a organização e estruturação de dados de forma padronizada e interoperável.

Alguns exemplos de aplicação do XML em bibliotecas são a criação de formatos de metadados como o *Dublin Core Metadata Initiative* (DCMI), que utiliza a sintaxe do XML, e a utilização de linguagens de programação que trabalham com XML para a criação de sistemas de gerenciamento de bibliotecas e repositórios digitais.

### **c) Resource Description Framework (RDF)**

*Resource Description Framework* (RDF) é uma linguagem de descrição de recursos, ou seja, uma linguagem de marcação para representar informações sobre recursos na *Web*, tais como livros, artigos, imagens, vídeos, entre outros. Essa linguagem é baseada em triplas RDF, que consistem em um sujeito, um predicado e um objeto, e

é projetada para ser interpretada por máquinas, tornando possível a troca e integração de dados entre diferentes sistemas (FERREIRA; SANTOS, 2013).

O RDF é uma linguagem de marcação voltada para a descrição de recursos e metadados, sendo utilizada principalmente em sistemas de recuperação de informações, *Web Semântica* e bibliotecas digitais. No contexto do trabalho do bibliotecário, o RDF pode ser utilizado para a descrição e recuperação de recursos de informação na *Web*, possibilitando a criação de sistemas mais precisos e eficientes para a busca e organização de informações.

#### ***d) Web Ontology Language (OWL)***

De acordo com McGuinness e van Harmelen (2004), *Web Ontology Language (OWL)* é uma linguagem de marcação baseada em XML para a descrição e definição de ontologias. Ontologias são estruturas conceituais que representam o conhecimento em uma determinada área, definindo classes, propriedades e relações entre conceitos.

O OWL é uma linguagem padronizada pela W3C (*World Wide Web Consortium*), juntamente com outras linguagens semânticas, como RDF e RDFS. A utilização do OWL em sistemas de informação e recuperação de informações permite uma melhor organização, busca e integração de dados e recursos em diferentes domínios e aplicações.

A linguagem possui três sublinguagens, OWL Full, OWL DL e OWL Lite, que diferem em suas capacidades expressivas e em suas restrições formais. A sublinguagem OWL Full é a mais expressiva, permitindo a definição de ontologias com recursos ilimitados, mas também a mais difícil de ser computacionalmente processada. A sublinguagem OWL DL é uma versão mais restrita, que mantém a expressividade semântica da OWL Full, mas com restrições formais que permitem uma maior eficiência no processamento computacional. Já a sublinguagem OWL Lite é a mais restrita, limitando-se a um subconjunto simples de construções da linguagem.

O OWL permite a definição de conceitos, propriedades e axiomas, permitindo a construção de ontologias complexas e detalhadas para representar conhecimento em uma determinada área de domínio. A linguagem é amplamente utilizada em

aplicações como a *Web Semântica*, recuperação de informações, gerenciamento de conhecimento, entre outras.

O papel do bibliotecário, como profissional da informação, está intrinsecamente ligado à organização, recuperação e disseminação de informações, seja em ambientes físicos ou virtuais. As linguagens de marcação, como HTML, XML, OWL e RDF, podem ser ferramentas importantes para auxiliar o bibliotecário no desempenho dessas atividades. O bibliotecário deve ter conhecimento dessas linguagens para conseguir desenvolver sistemas de organização e recuperação da informação mais eficazes.

### 3.6 Tecnologias e ferramentas para representação

A modelagem conceitual permite a representação e compreensão dos conceitos e relacionamentos presentes em um domínio específico. Para auxiliar nesse processo, diversas tecnologias e ferramentas estão disponíveis, oferecendo recursos para a representação visual, criação e edição de modelos conceituais. A escolha da ferramenta mais adequada dependerá das necessidades específicas do projeto, bem como das habilidades e recursos disponíveis da equipe de desenvolvimento. Existem diversas ferramentas disponíveis no mercado para auxiliar na modelagem de domínios. Algumas delas incluem:

- Protégé<sup>9</sup>: é uma ferramenta de código aberto para criação e edição de ontologias. Ela permite a criação de ontologias em OWL e RDF, além de oferecer recursos para inferência, validação e análise de ontologias. Esta ferramenta está disponível na versão desktop e na versão web;
  - WebProtégé<sup>10</sup>: é uma versão baseada em web do Protégé, com recursos para colaboração e compartilhamento de ontologias entre membros de uma equipe de desenvolvimento;
- TopBraid<sup>11</sup> Composer: é uma ferramenta de modelagem de ontologia que suporta as linguagens de modelagem OWL, RDF, RDFS e SKOS. Ela oferece recursos para criação e edição de ontologias, além de permitir a integração com outras ferramentas de modelagem e gerenciamento de dados;

---

<sup>9</sup> <https://protege.stanford.edu/>

<sup>10</sup> <https://webprotege.stanford.edu/>

<sup>11</sup> <https://www.topquadrant.com/>

- IBM Watson Studio<sup>12</sup>: é uma plataforma de análise de dados que inclui recursos para modelagem de dados e criação de ontologias. Ela permite a integração com outras ferramentas de análise de dados, como o Apache Spark, e oferece recursos para colaboração e compartilhamento de projetos;
- PowerDesigner<sup>13</sup>: é uma ferramenta de modelagem de dados que permite a criação de modelos conceituais, lógicos e físicos. Ela suporta diversas linguagens de modelagem, incluindo UML e BPMN, e oferece recursos para integração com outras ferramentas de gerenciamento de dados;
- ER/Studio<sup>14</sup>: é uma ferramenta de modelagem de dados que permite a criação de modelos conceituais, lógicos e físicos. Ela oferece recursos para gerenciamento de dados, incluindo a criação de dicionários de dados e a integração com outras ferramentas de gerenciamento de dados;
- Thesa<sup>15</sup>: é um software que permite a elaboração de múltiplos tesouros por multiusuários. Sua metodologia é baseada nas normas ISO e NIZO vigentes compatibilizadas com os requisitos semânticos presentes nas novas demandas dos sistemas de organização do conhecimento. Ele funciona em ambiente *Web*, pode ser baixado gratuitamente, e por se tratar de um open source (código aberto), pode ser modificado ou aperfeiçoado;
- PoolParty<sup>16</sup>: é uma plataforma de tecnologia semântica desenvolvida e licenciada pela *Semantic Web Company*. Ela utiliza tecnologias baseadas em padrões definidos pelo W3C e algoritmos de mineração de texto e processamento de linguagem natural (NLP) para extrair automaticamente termos relevantes de documentos para a construção de taxonomias, tesouros, ontologias e grafos de conhecimento empresarial;
- Onto4All<sup>17</sup>: é um editor de ontologias desenvolvido como parte do projeto OntoForInfoScience, que tem como objetivo facilitar a construção de ontologias por especialistas do domínio e cientistas da informação. É um editor gráfico de ontologias gratuito que permite criar, editar e exportar ontologias de forma fácil e intuitiva. Ele oferece uma interface gráfica para modelagem de ontologias, uma guia de regras de construção ontológica e uma extensa paleta de classes e relacionamentos ontológicos para facilitar a criação de ontologias.

---

<sup>12</sup><https://www.ibm.com/br-pt/cloud/watson-studio>

<sup>13</sup><https://www.sap.com/brazil/products/technology-platform/powerdesigner-data-modeling-tools.html>

<sup>14</sup><https://www.idera.com/products/er-studio/enterprise-data-modeling/>

<sup>15</sup> <https://www.ufrgs.br/tesouros/index.php/thesa>

<sup>16</sup> <https://www.poolparty.biz/>

<sup>17</sup><https://onto4all.com/pt#about>



### 3.7 Interdisciplinaridade com a linguística, a lógica e a semiótica

A modelagem conceitual, como uma área de estudo que busca representar conceitos e suas inter-relações, está intrinsecamente relacionada a várias disciplinas, como a linguística, a lógica e a semiótica. Essas disciplinas contribuem de maneiras diferentes para a compreensão e representação dos conceitos em um domínio específico e fornecem abordagens complementares que ajudam a estruturar, compreender e comunicar de maneira eficaz o conhecimento contido na informação.

#### **a) Linguística:**

A escolha e definição corretas dos termos são essenciais para a organização da informação. A linguística desempenha um papel fundamental na modelagem conceitual, uma vez que os conceitos são expressos por meio de linguagem. A análise linguística pode ajudar a identificar os termos e significados relevantes, além de fornecer uma base para a criação de uma terminologia consistente e compreensível. São levantadas questões sobre a adequação dos termos utilizados, sua precisão e a consistência na sua utilização. A linguística também aborda questões de semântica, sintaxe e pragmática, que são importantes para a representação precisa dos conceitos. A linguística analisa as relações de significado entre os elementos da informação. Essas relações incluem sinonímia, antonímia, hiperonímia, hiponímia e outras relações semânticas importantes para a organização adequada da informação.

#### **b) Lógica**

A lógica é de fato um ramo da filosofia que estuda as formas do pensamento, as operações intelectuais e o raciocínio. Ela está preocupada em determinar a validade e a consistência dos argumentos, bem como em estabelecer as relações entre as proposições. O raciocínio dedutivo, indutivo e abduutivo são importantes métodos lógicos utilizados para inferir conclusões a partir de premissas. Essas operações lógicas são fundamentais para a organização e representação da informação, pois permitem a estruturação coerente do conhecimento e a validação dos argumentos utilizados.

A lógica fornece uma estrutura formal para a definição de termos, a especificação de relações e a construção de inferências lógicas. Através de linguagens formais, como a lógica de predicados ou a lógica de descrição, é possível representar de forma precisa e sem ambiguidades os conceitos e suas relações. A lógica formal trata dos princípios de inferência e raciocínio válidos. Isso permite verificar a consistência e a validade lógica das informações representadas, bem como realizar inferências lógicas para derivar novas informações com base nas relações estabelecidas. Quando a comparação entre as características dos conceitos revela que dois conceitos diferentes possuem uma ou duas características em comum, surgem as relações entre esses conceitos. Existem diferentes tipos de relações que podem ser estabelecidas:

- **Relações lógicas:** envolvem relações de identidade, implicação, interseção, disjunção e negação. Essas relações se baseiam em princípios lógicos e são usadas para estabelecer conexões entre os conceitos.
- **Relações hierárquicas:** referem-se às relações de gênero e espécie, ou seja, conceitos mais amplos e conceitos mais restritos. Essas relações indicam uma relação de inclusão, em que um conceito mais amplo abrange um conceito mais específico.
- **Relações partitivas:** envolvem o todo e suas partes. Um conceito pode ser dividido em partes menores, e essa relação de parte-todo é estabelecida.
- **Relação de oposição:** envolve relações de contradição e contrariedade. Dois conceitos podem ser opostos em termos de suas características, sendo que a contradição implica em características mutuamente exclusivas, enquanto a contrariedade implica em características que não podem coexistir.
- **Relações funcionais:** aplicam-se a conceitos que expressam processos ou funções. Essas relações descrevem como um conceito está relacionado a outro no contexto de uma função ou atividade específica.

Na modelagem conceitual, a lógica desempenha um papel fundamental na representação e na definição de relações entre os conceitos. Essas relações entre conceitos são importantes para a organização e representação da informação, pois permitem estabelecer conexões semânticas entre os conceitos e facilitam a compreensão e recuperação da informação de forma mais estruturada.

### ***c) Semiótica***

A semiótica, por sua vez, estuda os signos e os processos de significação. Na modelagem conceitual, a semiótica desempenha um papel importante na

compreensão dos significados atribuídos aos conceitos e na interpretação dos modelos construídos. A semiótica ajuda a analisar os signos utilizados na representação dos conceitos, incluindo símbolos, ícones e índices, e como eles são interpretados pelos usuários.

#### **4 COMPETÊNCIAS E HABILIDADES NECESSÁRIAS AOS PROFISSIONAIS DE INFORMAÇÃO**

De acordo com Saracevic (1996), o papel da CI é fornecer os meios para o fornecimento de informações relevantes para indivíduos, grupos e organizações envolvidas com a ciência e tecnologia. Isso destaca a importância da CI em facilitar o acesso e o uso efetivo da informação para apoiar a tomada de decisões, a pesquisa científica, o avanço tecnológico e o progresso em diversas áreas do conhecimento. A CI tem como objetivo principal compreender os processos de produção, organização, disseminação e uso da informação, bem como os sistemas e tecnologias envolvidos nesses processos. Ela também busca investigar as necessidades de informação dos usuários e desenvolver métodos e ferramentas para melhorar a recuperação e a utilização da informação.

O profissional da informação, em qualquer contexto, deve desenvolver competências básicas em quatro categorias, segundo Valentim (2002), a saber: Competências de Comunicação e Expressão, Competências Técnico-Científicas, Competências Gerenciais e Competências Sociais e Políticas. Esse quarteto de competências básicas dos profissionais da informação, em particular do bibliotecário, demonstra como seu trabalho sempre esteve baseado em competências que hoje, mais e mais, são imprescindíveis ao tratamento da informação nos ambientes cada vez mais desafiadores pelo volume de dados e informações, por um lado; e por outro, pelo uso de ferramentas tecnológicas cada vez mais sofisticadas. No primeiro conjunto de competências se incluem “formular e gerenciar projetos de informação” e “executar procedimentos automatizados” (p. 123). Já nas competências técnico-científicas estão as competências mais diretamente relacionadas ao tratamento de informações e dados, desde coletar, armazenar, organizar, até proteger e manter os dados, dentro do escopo de “Desenvolver e executar o processamento de documentos em distintos suportes em unidades, sistemas e serviços de informação” (p. 123). No terceiro grupo estão as

competências gerenciais, que visam o controle de todos os processos, desde os mais procedimentais até “planejar, constituir e manipular redes globais de informação” (p. 124). Por último, mas não menos importantes estão as competências sociais e políticas, que envolvem ter “uma atitude crítica e criativa a respeito das resoluções de problemas e questões de informação” (p. 125). Assim, fica claro que os conhecimentos, habilidades e atitudes que o profissional da informação deve possuir e desenvolver para enfrentar os desafios do *Big Data* já estavam elencados dentre as competências típicas do profissional, já no início do séc. XXI.

Em outro documento, de atualização da base de conhecimento dos profissionais da informação, de 2021, são estabelecidos os princípios fundamentais de conhecimento e habilidades da expertise profissional, baseada em ética e valores; conceito mais amplo de biblioteca, dados, informações e contexto do setor de conhecimento; e contexto organizacional e ambiental. Os grupos de conhecimentos e habilidades incluem as seguintes categorias: expertise profissional e habilidades genéricas (CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS CILIP, 2021).

Dentre as competências mais duras, mais típicas da expertise do profissional da informação estão: 1) gestão da informação, que inclui desde a seleção dos fontes de informação, passando pelo uso das linguagens (classificação, taxonomias, ontologias, tesouros) até a arquitetura da informação (projeto, gerenciamento, metadados); 2) gestão do conhecimento, que envolve a gestão de conhecimento, informação e dados, incluindo síntese, transferência e compartilhamento em prol dos ativos e da própria organização; 3) exploração e uso da informação, envolvendo comportamento, serviços e recuperação da informação além de habilidades em análise, avaliação e síntese (resumos) da informação, utilizando a informetria; 4) pesquisa, que inclui desde o reconhecimento do valor da pesquisa, passando por tipos de pesquisa (inclusive métricas), contextos da pesquisa, comunicação e ética; 5) governança e conformidade (*compliance*) da informação, com as questões legais, desde o direito à informação, passando por direitos autorais, propriedade intelectual e licenciamento, até o compartilhamento, segurança e auditoria; 6) gerenciamento e arquivamento dos registros, do armazenamento ao descarte, envolvendo acesso, digitalização, curadoria, armazenamento e preservação; 7) gestão e

desenvolvimento de coleções, envolvendo seleção, avaliação e promoção; 8) Literacias (competências informacionais) e aprendizagem, desde a aprendizagem de indivíduos a comunidades, incluindo habilidades educacionais e a preocupação com ambientes de aprendizagem (CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS CILIP, 2021).

Já dentre as habilidades consideradas genéricas, não propriamente típicas do profissional da informação, estão: 1) foco nos utilizadores, design e marketing do serviço; 2) liderança, advocacy, influência e eficácia pessoal; 3) estratégia, planejamento e gestão; e 4) foco no cliente, *design* de serviço e *marketing*; e 5) tecnologia e comunicação, o que não envolve apenas a exploração dos recursos tecnológicos, mas também a utilização de habilidades de comunicação, de ferramentas colaborativas, de relações públicas e de competências linguísticas (CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS CILIP, 2021). Mais uma vez se confirma o vínculo e a importância do profissional da informação nas atividades da chamada Era do *Big Data*.

#### 4.1 O profissional da informação e o Big Data

Coneglian, Gonzalez e Santarém (2017) apresentam os principais desafios do *Big Data*, como volume, variedade, velocidade, veracidade e valor dos dados, e destacam a importância do profissional da informação para lidar com essas questões. Além disso, os autores abordam as habilidades e competências que os profissionais da informação devem possuir para atuar na área de *Big Data*, como o conhecimento de técnicas de mineração de dados, a capacidade de lidar com grandes volumes de informações, a habilidade de analisar e interpretar dados, e a capacidade de trabalhar em equipe e colaborar com outras áreas. Ainda é discutida a importância da formação e capacitação dos profissionais da informação para atuar no contexto do *Big Data*, destacando a necessidade de atualização constante e de formação de equipes multidisciplinares para lidar com os desafios da área.

Garritano e Carlson (2009) chamam a atenção para o papel das pessoas bibliotecárias especialistas em ciências e sua colaboração em projetos de e-ciência (*e-science*). Algumas das principais contribuições e conceitos do texto incluem suas habilidades na gestão da informação, incluindo organização, armazenamento,

preservação e compartilhamento. Os autores ainda consideram as habilidades para compreender as necessidades dos usuários como um ponto favorável. Por fim, estes autores destacam o papel destes profissionais para a identificação de fontes e recursos de informação.

Os autores Melo e Rochemback (2019) ressaltam que a formação em Arquivologia e Ciência da Informação é fundamental para os profissionais da informação, uma vez que essas áreas possuem expertise na organização, classificação e recuperação de informações, competências essenciais para lidar com o *Big Data*.

Ainda temos a contribuição de Reis e Fonseca (2020) que abordam a importância do profissional bibliotecário no contexto do *Big Data*, apresentando conceitos e desafios relacionados à coleta, processamento e análise de grandes volumes de dados. Estes autores discutem as principais habilidades e competências que uma pessoa bibliotecária deve ter para atuar nessa área, tais como o conhecimento em tecnologias de informação, a capacidade de gerir projetos e equipes, além da habilidade de análise de dados.

Souza et al. (2013) discutem as mudanças que ocorrem na área de CI em decorrência da revolução tecnológica, destacando quatro tópicos que têm influenciado o campo: *Big Data*, *Cloud Computing*, Redes Sociais e *Web Semântica*. O conceito de *Big Data* é definido pelos autores como um conjunto de tecnologias e técnicas que permitem o armazenamento, o processamento e a análise de grandes quantidades de dados, estruturados e não estruturados, de diferentes fontes e formatos, em tempo real ou quase real. Além disso, é ressaltado que o *Big Data* representa um desafio e uma oportunidade para a CI, que pode contribuir com seu conhecimento e experiência na organização, recuperação, análise e uso da informação. Eles enfatizam a importância da CI em acompanhar e adaptar-se a essas mudanças tecnológicas para continuar sendo relevante na sociedade.

Dziekaniak (2004) discute o papel do bibliotecário na criação de projetos de software e seu envolvimento com a tecnologia da informação. O autor enfatiza a importância do bibliotecário na equipe de desenvolvimento de software, enfatizando sua habilidade em gerenciar a informação e na identificação das necessidades dos usuários. O artigo aborda ainda as competências necessárias ao

bibliotecário para atuar nesse ambiente e as possíveis contribuições do profissional para o desenvolvimento de projetos de software mais eficazes e alinhados às necessidades dos usuários. O profissional de ciência da informação pode contribuir com conhecimentos específicos para a criação de bancos de dados e oferecer uma perspectiva técnica para indexação e recuperação de informação (DZEKANIAK, 2004).

Por fim, Rodrigues, Nóbrega, e Dias (2017) investigam os desafios da gestão de dados na era do *Big Data* e como isso afeta a sociedade contemporânea, especialmente no campo da CI. Os autores discutem o perfil profissional do cientista de dados e as habilidades necessárias para lidar com um ambiente de dados em grande escala, explorando a compreensão do que são esses dados, como utilizá-los estrategicamente e a complexidade envolvida na gestão de dados. Ainda abordam as potencialidades, riscos e oportunidades na sociedade contemporânea e destaca a importância de desenvolver um novo perfil profissional na área de gestão de dados, ressaltando a necessidade de ampliar as pesquisas sobre o assunto dentro do campo da CI.

#### *4.1.1 Biblioteconomia de Dados*

A Biblioteconomia de Dados é um campo de estudo que envolve bibliotecários no gerenciamento e curadoria de dados de pesquisa em várias disciplinas científicas. Essa área busca fornecer técnicas e serviços para apoiar a organização, armazenamento, descoberta e acesso aos dados de pesquisa. A Biblioteconomia de Dados surgiu com o crescimento das coleções de dados de pesquisa, especialmente os dados quantitativos nas ciências sociais, e o estabelecimento de arquivos e bibliotecas de dados (SEMELER; PINTO, 2019).

O bibliotecário de dados desempenha um papel crucial na garantia da integridade, qualidade e acessibilidade dos dados utilizados no ciclo da produção do conhecimento científico. O profissional de biblioteconomia com habilidades em gestão de dados científicos é capaz de lidar com diferentes etapas desse ciclo, desde a coleta e armazenamento dos dados até sua análise e disseminação. Ele possui conhecimentos sobre esquemas de metadados, padronização de dados, políticas de compartilhamento e preservação de dados, além de compreender as necessidades



dos pesquisadores no que diz respeito à gestão e acesso aos dados (SEMELER; PINTO, 2019).

Além de fornecer suporte na descoberta e acesso a recursos de dados, os bibliotecários de dados podem oferecer consultoria sobre o tratamento e a aquisição de dados para análise de pesquisa em disciplinas específicas. Eles desempenham um papel importante na interpretação, tomada de decisão e comunicação de dados, além de estarem envolvidos na preservação e avaliação de valores econômicos atribuídos aos dados.

Ao trazer neste trabalho a reflexão sobre o bibliotecário de dados, destaca-se no trabalho de Semeler e Pinto (2019) sua contribuição sobre a demanda por habilidades como capacidade de avaliação e coleta de dados, promoção e *marketing* de dados, coordenação de práticas com dados e negociação de expectativas de dados. Para os autores, este profissional também deve ter habilidades tecnológicas e conhecimentos em programação, manipulação de dados, metadados e tecnologias relacionadas. As habilidades necessárias para os bibliotecários de dados incluem habilidades interpessoais, conhecimento contextual sobre ambientes institucionais, uso de dados e conhecimentos sobre tecnologias de manipulação de dados. Os bibliotecários de dados devem combinar habilidades técnicas com habilidades tradicionais de bibliotecários, como catalogação, desenvolvimento de coleções e serviços de referência.

#### 4.2 Competências e Habilidades

Competências técnicas e comportamentais necessárias para um profissional da informação podem variar, dependendo da área de atuação e das necessidades específicas do profissional. Ao abordar como o profissional da informação pode atuar no âmbito do *Big Data* e da modelagem conceitual ou de domínio pode-se destacar a necessidade de desenvolver competências e habilidades que são apresentadas a seguir.

As competências técnicas estão relacionadas a ação dos profissionais a respeito à capacidade de lidar com os conteúdos - conceitos, comportamentos e atitudes, e à

habilidade de reconstruí-los. Se refere à habilidade de realização de uma determinada tarefa (VITORINO; PIANTOLA, 2020).

Tendo em vista a literatura estudada, podemos demarcar um conjunto de competências técnicas - não restrita a elas - que são desejáveis ao profissional da informação para atuar com a modelagem de domínio, são elas:

- Familiaridade com técnicas de modelagem conceitual: O profissional da informação deve ter conhecimento e habilidades em técnicas de modelagem conceitual, como diagramas de entidade-relacionamento (ER), diagramas UML, sistemas de vocabulários controlados, taxonomias, tesouros e ontologias, entre outros. Ao ter familiaridade com essas técnicas de modelagem conceitual, o profissional da informação será capaz de representar de forma precisa e clara as entidades, atributos e relacionamentos do domínio em estudo, garantindo uma base sólida para a organização, recuperação e compartilhamento de informações. Além disso, essa habilidade permite uma melhor comunicação com especialistas e usuários do domínio, facilitando a colaboração e o desenvolvimento de soluções eficazes.
- Conhecimento de ontologias e metodologias para construção de ontologias: O profissional da informação deve ter conhecimento em ontologias, que são modelos de conhecimento que representam conceitos, propriedades e relacionamentos em um determinado domínio. Ao possuir conhecimento em ontologias, o profissional da informação é capaz de criar e utilizar essas estruturas de representação do conhecimento para modelar e organizar informações dentro de um domínio específico. Isso permite a criação de modelos conceituais mais robustos, precisos e interoperáveis.
- Habilidade em aquisição de conhecimento: A modelagem de domínios envolve técnicas de aquisição e análise de conhecimento com intuito de melhor conhecer o domínio que se deseja modelar. É fundamental ter um bom entendimento do domínio de conhecimento no qual o modelo conceitual será desenvolvido.
- Conhecimento de técnicas de aquisição de conhecimento: Existem várias técnicas e métodos de aquisição de conhecimento que podem ser utilizados para obter informações e compreender um determinado domínio ou problema, a subseção 4.2.1 sintetiza algumas destas técnicas e métodos.
- Habilidades na definição dos termos e conceitos: A formulação de definições de termos é aplicada em sistemas de terminologia, classificações, tesouros e ontologias, entre outros. Uma boa definição dos termos e conceitos garante compreensão clara e consistente dos termos e conceitos utilizados no domínio, facilitando a comunicação, a interpretação correta das informações,

minimizando ambiguidades e promovendo o desenvolvimento de modelos conceituais precisos e bem fundamentados.

- Familiaridade com lógica e lógica formal: Esta é uma competência técnica importante para profissionais envolvidos na modelagem conceitual. A lógica formal é uma disciplina que se concentra na estrutura e na validade dos argumentos, fornecendo uma base sólida para a representação formal dos conceitos e relações em um modelo conceitual. Ao aplicar a lógica formal na modelagem conceitual, é possível estabelecer axiomas, regras e restrições que governam o comportamento dos conceitos, e das relações entre os conceitos, garantindo a consistência e a coerência do modelo.
- Habilidade em trabalhar com ferramentas de modelagem: O profissional da informação deve ter habilidade em trabalhar com ferramentas de modelagem, como Protégé, TopBraid Composer, OntoGraf, Thesa, entre outras, e capacidade de utilizá-las para criar ontologias e modelos de conhecimento. Ao ter habilidade em trabalhar com essas ferramentas, o profissional da informação é capaz de aproveitar ao máximo as funcionalidades oferecidas por elas. Isso inclui recursos como a criação de classes e instâncias, definição de propriedades e relacionamentos, especificação de restrições e axiomas, e visualização de diagramas conceituais.
- Conhecimento de linguagens de marcação: O profissional da informação deve ter conhecimento em linguagens de modelagem, como XML, UML (Unified Modeling Language), RDF, OWL, entre outras; As linguagens de marcação fornecem um conjunto de regras e convenções para estruturar e descrever informações de forma organizada e padronizada. Ao utilizar linguagens de marcação adequadas, o profissional da informação é capaz de representar de maneira precisa e semântica os conceitos, relacionamentos e características do domínio em um modelo conceitual.
- Conhecimento de técnicas para estabelecer categorias: O profissional da informação deve ter habilidade para classificar e organizar informações em categorias, como áreas, subáreas e assuntos. Essas técnicas ajudam a criar uma estrutura conceitual coerente e facilitam a compreensão e navegação pelos dados no modelo de domínio. Essas técnicas permitem classificar e organizar as informações em categorias adequadas, como áreas, subáreas e assuntos, proporcionando uma estrutura conceitual coerente e facilitando a compreensão e navegação pelos dados no modelo de domínio.
- Conhecimento sobre metadados e padrões de metadados: Os padrões de metadados, como o *Dublin Core*, por exemplo, fornecem um conjunto de elementos padronizados que podem ser utilizados para descrever os recursos. Esses elementos podem ser mapeados para os termos e propriedades definidos no modelo, proporcionando uma conexão entre os metadados e os conceitos representados. Isso permite uma melhor integração e interoperabilidade dos dados, facilitando a recuperação e o

compartilhamento das informações. Além disso, as anotações semânticas desempenham um papel importante no enriquecimento dos dados ao adicionar significado adicional por meio de metadados estruturados. As anotações semânticas podem incluir informações como termos controlados, relações semânticas, classificações e outros elementos que enriquecem a representação dos dados. Essas anotações podem ser baseadas em padrões de metadados e ontologias para garantir a consistência e a precisão na descrição dos recursos.

- **Habilidades de comunicação e colaboração:** Capacidade de comunicar efetivamente com especialistas e usuários do domínio, trabalhar em equipe na construção do modelo conceitual e colaborar com outros profissionais envolvidos no projeto.

Por fim, resgatando ainda Sayão (2001) que contribui ao discutir que a mente humana busca decompor o mundo real em sistemas simplificados, permitindo obter uma visão das características essenciais de um domínio específico. Essa simplificação requer criatividade tanto sensorial quanto intelectual, o que implica reconhecer que, ao construir modelos, algumas características da realidade que não estão diretamente relacionadas aos objetivos pretendidos são ignoradas ou deixadas de lado em prol de uma maior inteligibilidade ou facilidade de compreensão.

De acordo com Ribeiro (2008) o processo de organização de dados e informações precisa de mais instrumentação, pois atualmente há uma maior ênfase nos aspectos tecnológicos em detrimento das questões relacionadas à gestão da informação, deixando em segundo plano as reflexões sobre a organização das informações.

Ao desenvolver essas habilidades, o profissional estará melhor preparado para enfrentar os desafios da modelagem conceitual e garantir a qualidade e eficiência do resultado final.

#### *4.2.1 Técnicas e Métodos de Aquisição de Conhecimento*

Sobre técnicas e métodos de aquisição de conhecimento existem diversos trabalhos que discutem o assunto, cita-se Almeida e Coelho (2012) que apresentam um roteiro metodológico para a aquisição de conhecimento voltado à construção de ontologias no domínio da hematologia. O objetivo do estudo é fornecer uma abordagem sistemática e estruturada para a obtenção de conhecimento relevante e necessário

para a criação de ontologias, que são estruturas de representação do conhecimento que buscam capturar as relações entre conceitos em um determinado domínio. No roteiro metodológico proposto pelos autores eles apontam algumas técnicas como

- Levantamento bibliográfico: Realiza-se uma revisão da literatura existente no domínio da hematologia, buscando identificar fontes confiáveis de informação e estabelecer uma base sólida de conhecimento prévio.
- Identificação de especialistas: Identificam-se especialistas no domínio da hematologia que possam contribuir com conhecimento especializado e experiência prática.
- Entrevistas com especialistas: São realizadas entrevistas estruturadas com os especialistas identificados, com o objetivo de extrair informações relevantes sobre o domínio, conceitos, relações, características e restrições.

O estudo de Farinelli (2017), por sua vez, relata o uso de diversas técnicas e métodos na experiência de aquisição de conhecimento para a construção da ontologia OntONeo. Algumas dessas técnicas incluem:

- Entrevistas: Realizar entrevistas com especialistas, profissionais ou usuários experientes no domínio em questão. As entrevistas permitem obter insights valiosos, conhecimento especializado e informações detalhadas sobre o domínio de interesse.
- Observação: Observar diretamente pessoas, processos ou eventos relevantes para o domínio. A observação pode ajudar a identificar padrões, comportamentos e problemas que não seriam facilmente capturados por outros métodos.
- Questionários e pesquisas: Elaborar questionários estruturados ou realizar pesquisas para coletar informações quantitativas ou qualitativas. Esses métodos são úteis para obter uma visão ampla e generalizada sobre o domínio.
- Revisão de literatura e documentos: Analisar livros, documentos, relatórios, manuais, artigos científicos, entre outros materiais relevantes para obter conhecimento existente sobre o domínio em questão. A revisão de literatura e documental pode ajudar a compreender teorias, conceitos e práticas estabelecidas.
- Grupos focais: Realizar sessões de discussão em grupo com especialistas ou usuários representativos do domínio. Os grupos focais permitem obter diferentes perspectivas, explorar ideias e identificar desafios ou oportunidades relevantes.

- Mineração de dados: Utilizar técnicas de mineração de dados para extrair informações valiosas a partir de grandes conjuntos de dados. Essas técnicas podem ajudar a descobrir padrões, tendências e relações ocultas no domínio.
- Prototipagem rápida: Criar protótipos ou modelos iniciais que representem as principais características do domínio. Através da construção e teste de protótipos, é possível explorar e validar o conhecimento adquirido de forma mais tangível.

Cada técnica tem suas próprias vantagens e contribui para uma compreensão mais completa do domínio em questão. A combinação dessas técnicas permite abordar diferentes aspectos do conhecimento, como conhecimento especializado, conhecimento prático, informações quantitativas e qualitativas, conhecimento existente, perspectivas múltiplas e *insights* derivados de dados. Ao utilizar esses métodos de forma integrada, os profissionais podem adquirir uma visão mais ampla e precisa do domínio, capturando os conceitos, relações e restrições relevantes para a construção do modelo conceitual. É importante destacar que a escolha das técnicas e métodos de aquisição de conhecimento depende do contexto, dos recursos disponíveis e dos objetivos específicos do projeto. Cada domínio pode exigir abordagens diferentes e adaptadas às suas características particulares.

#### *4.2.2 Formação em Modelagem*

A formação necessária para um profissional da informação que trabalha com modelagem de domínios seja em ambientes tradicionais ou de *Big Data* podem ser obtidos por meio de diferentes abordagens acadêmicas e práticas. Embora não haja um caminho único e definitivo, existem algumas opções comumente consideradas para adquirir as competências e habilidades necessárias.

No caso de formação acadêmica, esta pode ser multidisciplinar, sugere-se uma base de cursos de graduação e pós-graduação que podem preparar estes profissionais, por exemplo:

- Graduação em Ciência da Informação, Biblioteconomia, Arquivologia, Ciência da Computação, Engenharia de Computação, Análise de Sistemas ou áreas afins. Esses cursos fornecem uma base sólida em modelagem da informação, gestão da informação, recuperação de informação e tecnologias da informação, que são fundamentais para entender os princípios e conceitos

subjacentes à modelagem de domínio e ferramentas que podem ser usadas para este fim.

- Pós-graduação em Ciência da Informação, Gestão da Informação, Organização da Informação, Engenharia de Software, Sistemas de Informação, Ciência de Dados, Big Data, Banco de Dados, entre outros. Os programas de pós-graduação oferecem oportunidades para aprofundar o conhecimento em áreas específicas relacionadas à modelagem de domínio. Alguns oferecem abordagens mais direcionadas à modelagem de bancos de dados e outros já oferecem abordagem originária a dados de sistemas de organização do conhecimento como tesauros e ontologias.

Complementando, existem no mercado atualmente treinamentos e aperfeiçoamentos mais específicos que seguem geralmente as tendências mercadológicas da atuação profissional. Em geral, são treinamentos mais voltados para ferramentas e metodologias e visam preparar o profissional para a prática, estes treinamentos envolvem as técnicas e as ferramentas mencionadas anteriormente na lista de competências e habilidades apresentada, nominalmente são:

- Treinamentos em técnicas de modelagem sobre técnicas de modelagem como diagramas ER, UML, taxonomias e ontologias e outros.
- Treinamentos em ferramentas de modelagem que visam demonstrar ferramentas de modelagem amplamente utilizadas, como Protégé, TopBraid Composer, OntoGraf e outras.
- Treinamentos em linguagens de marcação e modelagem, como XML, RDF e OWL.
- Treinamentos sobre os conceitos e práticas em Big Data e Ciência de Dados.

Além desses treinamentos e formações, é importante que o profissional da informação mantenha-se atualizado com as tendências e tecnologias em evolução na área de modelagem, representação e recuperação da informação, por meio de leitura de artigos científicos, participação em conferências e eventos, além de prática constante no mercado de trabalho. Existe uma vasta literatura disponível sobre os assuntos abordados neste artigo que também estão disponíveis e servem como boas referências para quem deseja se especializar.

Vale ressaltar que a formação é um processo contínuo. À medida que a tecnologia evolui e novas abordagens surgem, é fundamental que o profissional da informação



esteja disposto a se adaptar e buscar aprendizado constante para se manter atualizado e relevante no campo da modelagem de domínio.

## **5 EXEMPLOS DE APLICAÇÃO**

Nos últimos anos, a Era do *Big Data* alavancou o surgimento de novas aplicações para os conhecimentos da Ciência da Informação. Elencamos alguns exemplos dessas novas aplicações das práticas do profissional da informação na modelagem de domínio em ambientes não tradicionais.

### **5.1 Categorização, Classificação e Representação da Informação Aplicada ao E-commerce**

As categorias de produtos são uma forma de classificar e organizar as mercadorias de uma loja virtual, conforme os seus nichos de vendas e o perfil dos visitantes (identificar as necessidades desses usuários e considerá-las na modelagem do domínio).

A categorização de produtos requer primeiramente a sua representação conceitual, o que permitirá posteriormente agrupar os produtos semelhantes e separar os produtos diferentes, de acordo com alguma característica. Esse processo é realizado principalmente utilizando técnicas para a elaboração de taxonomia.

Segundo Ruscher (2022):

Basicamente, na classificação, organizamos as coisas dentro de “caixinhas”. E estas caixinhas são organizadas em uma estrutura hierárquica, como uma taxonomia. A taxonomia é uma classificação sistemática, em que as classes (caixinhas) se apresentam segundo uma ordem lógica, apoiada em princípios, contendo relações hierárquicas, ou seja, do mais geral para o mais específico. No marketplace, ela vai servir especialmente para organizar as categorias dos produtos na plataforma, através do sistema de navegação.

A estruturação lógica das páginas facilita a marcação semântica, a compreensão pelos buscadores e proporciona uma navegação mais fluida para os visitantes (melhora a experiência do usuário), o que pode aumentar a conversão das vendas.

Cavalcante e Bräscher (2014) afirmam:

Observa-se que, mais do que um componente da estrutura navegacional de um sítio, as taxonomias navegacionais estão intrinsecamente relacionadas à organização e recuperação de informação e podem ser preponderantes na decisão de compra de determinado produto ou serviço. Revela-se, então, a importância de que sejam capazes de atender às necessidades informacionais dos usuários, atuando como guias intuitivos durante a busca pelo item desejado.

A descrição física do produto é utilizada para representá-lo no ambiente digital, essa representação das informações do produto origina os metadados, que são dados estruturados e padronizados que descrevem as propriedades (atributos) do objeto informacional.

Na modelagem de domínio, o metadado pode representar uma entidade que possui conceito e relações com outras entidades, por exemplo, o metadado cor possui um conceito e relação com a entidade produto.

A aplicação destes princípios e instrumentos podem melhorar a qualidade dos dados dos produtos que é essencial para outros tratamentos que normalmente são realizados em empresas de tecnologia como *analytics*, algoritmos de recomendação e classificação, indexação automática etc (RUCHER, 2022).

## 5.2 Modelagem de Domínio e Organização da Informação Aplicada ao Desenvolvimento de *Chatbot s*

Os *chatbots* são programas que simulam uma conversa, com o objetivo de responder perguntas de tal forma que a pessoa que estiver interagindo com ele tenha a impressão de estar conversando com outra pessoa (AQUINO; ADANIYA, 2018).

*Chatbot s* são projetados nas empresas para interagir com os clientes, muitas vezes em linguagem natural utilizando inteligência artificial. O cliente expõe uma ou mais questões que deseja resolver (intenção) por um canal de comunicação, o *chatbot* analisa a questão e envia de volta uma resposta adequada para ele.

A análise e modelagem do domínio é essencial para essa interação com os *chatbots*, pois permite a compreensão da linguagem e das necessidades dos clientes.

Dias (2021) afirma que:

[...] para dizer o essencial, a Análise de Domínio explora, por uma perspectiva pragmática, uma comunidade discursiva e suas dimensões. O intuito é o de identificar conceitos, linguagens, ideologias, culturas e práticas compartilhadas por grupos específicos com a finalidade de dar contexto sociocultural e histórico à informação de determinado domínio e das comunidades que o constituem. Nada mais relevante do que “falar a língua” do grupo de usuários com o qual irá se comunicar.

A autora também afirma visualizar outras atividades da organização da informação que podem ser aplicadas no desenvolvimento de *chatbots* como: análise, interpretação, tradução, classificação, indexação e catalogação da informação.

Aprendendo sobre a criação de um *chatbot*, visualizei cada uma das atividades listadas sendo desenvolvidas em diferentes estágios da construção do assistente virtual. Desde a fase inicial de planejamento até a etapa de manutenção, a curadoria. Em uma das etapas, a de interação, na qual se elabora o desenho conversacional entre máquina e usuário, é necessário mapear e classificar as intenções dos usuários e descrevê-las por verbos. As intenções irão orientar o aprendizado da inteligência artificial. Mais adiante, na fase de curadoria, será preciso reclassificar essas intenções de acordo com respostas externas, distinguindo-as e agrupando-as por suas diferenças e similaridades (DIAS, 2021).

Artefatos informacionais utilizados na organização das informações como o vocabulário controlado, a taxonomia e o tesouro quando aplicados na construção, treinamento e curadoria dos *chatbots* contribuem para que o programa:

1. identifique as intenções dos clientes com mais precisão (controle de sinônimos), permitindo que o *chatbot* forneça uma resposta mais relevante e assertiva;
2. defina as respostas de maneira mais precisa e consistente (respostas organizadas de maneira hierárquica), evitando ambiguidades;
3. melhore a experiência do cliente, pois garanta que a interação com o *chatbot* seja clara e eficiente, aumentando a satisfação com esse canal oferecido pela empresa.

### 5.3 Vocabulário Controlado e Taxonomia Aplicados no SEO Semântico

Uma das dificuldades técnicas que ocorre quando um assunto é pesquisado na web é identificar o que o usuário perguntou em linguagem natural e combiná-lo com o conteúdo disponibilizado em sites, documentos e outros materiais. O buscador

precisa unir essas duas pontas da pesquisa, para isso precisa compreender a linguagem, entender o conceito das palavras utilizadas na busca e compreender o que o conteúdo desses materiais disponibilizados na web querem dizer.

Segundo Silva (2022), SEO Semântico é o processo de construir significados para o alinhamento de conceitos em conteúdos e sites na web. Através dessa construção, o profissional de SEO ajuda os buscadores semânticos a melhorar o entendimento sobre estes mesmos conteúdos, fazendo com que os algoritmos os qualifiquem como de alta qualidade, aumentando a sua presença nas páginas de resultados de buscas.

O profissional de SEO Semântico para conseguir tornar claro para os buscadores os conceitos presentes nos seus conteúdos (otimizar a semântica), precisa conhecer e modelar o domínio do conteúdo de seus sites e uma das maneiras de modelar esse domínio é por meio da criação de vocabulários controlados (lista de termos) e de taxonomias (relações hierárquicas entre os termos).

Uma das atividades mais complexas, de quando inicio um projeto de SEO Semântico é de entender os conceitos, e conseqüentemente os termos e entidades, do domínio do conhecimento ao qual o site pertence (SILVA, 2023).

Identificar e conceituar os termos e entidades de um domínio de um site colaboram para reduzir um dos principais obstáculos na recuperação dessa informação nas ferramentas de buscas na web que utilizam linguagem natural: a ambigüidade. Os termos e entidades contextualizam o site e conseqüentemente aumentam o grau de certeza no momento de posicioná-lo em uma página de resultado de uma pergunta que possui uma intenção de busca específica.

O sistema de recuperação ao encontrar diferentes significados possíveis de serem extraídos de uma frase ou palavra precisa distinguir um destes significados segundo o contexto, obtendo maior precisão na resposta dada ao usuário. As estratégias taxonômicas de organizar os conceitos segundo suas relações de significado, quando aplicadas ao sistema de recuperação possibilitam essa identificação do contexto de maneira mais precisa a partir de buscas anteriores do usuário (SILVA, 2022).

## 6 CONSIDERAÇÕES FINAIS

A modelagem conceitual é uma abordagem fundamental tanto na Ciência da Computação quanto na Ciência da informação. Ela consiste em representar de forma abstrata e semântica os conceitos e relacionamentos de um domínio de interesse, visando facilitar a compreensão e comunicação entre os envolvidos no desenvolvimento de sistemas de informação. Diversos trabalhos têm contribuído significativamente para o avanço da modelagem conceitual nas áreas de Ciência da Computação e Ciência da informação.

O estudo apresentou o contexto e importância da Ciência da Informação (CI) na organização, representação de modelagem de dados e informação e no âmbito da modelagem de domínio. Em resposta ao objetivo proposto foram descritas as principais técnicas, metodologias, práticas, tecnologias e instrumentos de representação necessários para modelagem de domínio.

No contexto de dados e informação, a CI desempenha um papel fundamental na compreensão dos aspectos teóricos e práticos relacionados à gestão de dados, organização da informação, recuperação da informação, preservação digital, análise de dados, entre outros tópicos relevantes. A CI também explora questões éticas, legais e sociais relacionadas à coleta, uso e compartilhamento de dados e informações.

Para a modelagem de domínios são necessárias habilidades específicas para compreensão do domínio, aquisição de conhecimento, e representação de dados. O desenvolvimento dessas habilidades é fundamental para garantir uma modelagem de domínios eficiente e de qualidade. O profissional da informação é o principal responsável pelo tratamento da informação, ele pode desempenhar um papel importante na discussão sobre o *Big Data*. Sua experiência em gestão da informação, organização de dados e busca de informações relevantes pode ser aplicada na análise e interpretação dos dados em grande escala. Além disso, eles podem ajudar a desenvolver estratégias e políticas para lidar com o volume crescente de dados e garantir sua qualidade, segurança e acessibilidade.

Neste artigo, exploramos as aplicações emergentes das práticas do profissional da informação na modelagem de domínio em ambientes não tradicionais. Com o

advento da Era do *Big Data*, as possibilidades de aplicação têm se expandido e se mostrado essenciais para lidar com a crescente quantidade de informações disponíveis. Nos exemplos apresentados, podemos observar como a categorização, classificação e representação da informação são cruciais para o sucesso do e-commerce. A organização dos produtos por meio de taxonomias facilita a navegação dos usuários, melhora a experiência do cliente e impulsiona a conversão de vendas.

Além disso, destacamos a importância da análise e modelagem do domínio no desenvolvimento de *chatbots*. Compreender a linguagem e as necessidades dos clientes é fundamental para criar interações mais naturais e eficientes. A aplicação de vocabulários controlados e taxonomias contribui para identificar com precisão as intenções dos usuários, melhorando a qualidade das respostas e a satisfação geral. No contexto do SEO Semântico, observamos como a modelagem do domínio por meio de vocabulários controlados e taxonomias pode aprimorar a compreensão dos buscadores em relação ao conteúdo dos sites. Isso resulta em uma melhor classificação nos resultados de busca e maior visibilidade online.

Esses exemplos reforçam a importância da Ciência da Informação na era digital. A habilidade de organizar, classificar e representar a informação de forma eficaz é essencial para lidar com a abundância de dados disponíveis atualmente. Através da aplicação das práticas da Ciência da Informação, podemos otimizar a recuperação da informação, melhorar a experiência do usuário e impulsionar os negócios.

Em suma, as práticas do profissional da informação desempenham um papel fundamental na modelagem de domínio em ambientes não tradicionais. O conhecimento e a aplicação adequada dessas práticas podem proporcionar benefícios significativos, como aumento das vendas, melhor interação com os clientes e maior visibilidade online. A constante evolução da Era do *Big Data* certamente trará novos desafios e oportunidades, e os profissionais da informação têm um papel fundamental a desempenhar na gestão e aproveitamento dessas informações em benefício das organizações e da sociedade como um todo.

Como perspectivas futuras de pesquisa sugere-se a realização de estudos de casos práticos da atuação do profissional da informação na modelagem de domínio relacionada ao catálogo de dados (organização dos dados e suas fontes em domínios,

descrição de metadados, linhagem e recuperação de dados), à curadoria de dados (governança e proteção de dados) e à UX (*User Experience*) Designer no contexto da organização e representação da informação.

## REFERÊNCIAS

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da informação**, Brasília, v. 31, p. 5-13, 2002.

ALMEIDA, M.B. **Ontologia em Ciência da Informação**. Teoria e Método. Curitiba:CRV, 2020.373p. (Representação do Conhecimento em Ciência da Informação, v.1)

ALMEIDA, M.B. **Ontologia em Ciência da Informação**. Tecnologia e Aplicações. Curitiba:CRV, 2021a.372p. (Representação do Conhecimento em Ciência da Informação, v.2)

ALMEIDA, M.B. Teorias ontológicas para modelagem. **Fronteiras da Representação do Conhecimento**, Belo Horizonte, 1, n. 2, p. 95-126, 2021b.

ALMEIDA, M.B.; COELHO, K.C. Aquisição de conhecimento para construção de ontologias: uma proposta de roteiro metodológico aplicado ao domínio da hematologia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. 35, p. 47-73, 2012.

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

AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. Alta Books Editora, 2016.

AQUINO, V.H.O.; ADANIYA, M.H.A.C. Desenvolvimento e aplicações de *chatbot*. **Revista Terra & Cultura: Cadernos de Ensino e Pesquisa**, [S. l.], v. 34, n. esp., p. 56-68, jul. 2018. Disponível em: <http://periodicos.unifil.br/index.php/Revistatestes/article/view/297>. Acesso em: 19 mar. 2023.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information retrieval: The concepts and technology behind search**. 2a ed. New York: Addison-Wesley Publishing Company, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, v. 284, n. 5, p. 34-43, 2001. Disponível em: <https://www.scientificamerican.com/article/the-semantic-web/>. Acesso em: 07 feb. 2023.

BORKO, H. Information science: what is it? **American documentation**, 19, n. 1, p. 3-5, 1968.



BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? *In: IX ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 2008. **Anais..** São Paulo, 2008. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/176535>. Acesso em: 20/03/2023.

BRAY, T.; PAOLI, J.; Sperberg-McQueen, C. M.; Maler, E.; Yergeau, F. **Extensible Markup Language (XML) 1.0** Fifth Edition. World Wide Web Consortium (W3C), 26 Nov. 2008. Disponível em: <https://www.w3.org/TR/xml/>. Acesso em: 19 mar. 2023.

CAMPOS, M. L. D. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, 33, p. 22-32, 2004.

CAMPOS, M. L. D. A.; GOMES, H. E. Princípios para modelagem de domínio: a posição de Barry Smith e de Ingetraut Dahlberg. **Ciência da Informação**, Brasília, 43, n. 1, 2014.

CAVALCANTE, Raphael da Silva; BRÄSCHER, Marisa. Taxonomias navegacionais em sítios de comércio eletrônico: critérios para avaliação. **Transinformação**, Campinas, v. 26, n. 2, maio-ago 2014. Disponível em: <https://doi.org/10.1590/0103-37862014000200008>. Acesso em: 02 abr. 2023.

CESARINO, M. A. N. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, 1985.

CHARTERED INSTITUTE OF LIBRARY AND INFORMATION PROFESSIONALS CILIP. **Professional Knowledge and Skills Base (PKSB)**. 2021. Disponível em [https://cdn.ymaws.com/www.cilip.org.uk/resource/resmgr/cilip/membership/benefits/pksb/pksb\\_old\\_to\\_new\\_mapping.pdf](https://cdn.ymaws.com/www.cilip.org.uk/resource/resmgr/cilip/membership/benefits/pksb/pksb_old_to_new_mapping.pdf). Acesso em 08 abr. 2023.

CHEN, P. P. **Modelagem de dados**: a abordagem entidade-relacionamento para projeto lógico. Makron Books do Brasil, 1990.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O Profissional da Informação na Era do Big Data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 22, n. 50, p. 128, 2017.

COREMANS, C. **HTML: A Beginner's Tutorial**. Brainy Software Inc, 2015.

COUGO, P. S. **Modelagem conceitual e projeto de banco de dados**. 1ª ed. 18ª Reimp. ed. Rio de Janeiro: Elsevier, 1997.

COX, M.; ELLSWORTH, D. English, Phoenix, AZ, USA. Application-controlled demand paging for out-of-core visualization. **IEEE Computer Society Press**. 1997, 235-ff. Disponível em: <https://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>. Acesso em: 18 de mar. 2023.

DAMA, Data Management Association International. **Data management body of knowledge 2.0 (DAMA-DMBOK 2)**. Bas King Ridge, New Jersey, USA: Technics Publications LLC, 2017. 624 p.

DAHLBERG, I. Teoria da classificação, ontem e hoje. In: **CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA**. 1972. p. 352-370.

DAHLBERG, I. Teoria do conceito. **Ciência da informação**, Brasília, v. 7, n. 2, 1978.

DIAS, B. *Chatbot s*: primeiras impressões de uma bibliotecária na criação de assistentes virtuais inteligentes. **LinkedIn Pulse**. out. 2021. Disponível em: <https://www.linkedin.com/pulse/chatbots-primeiras-impress%C3%B5es-de-uma-bibliotec%C3%A1ria-na-beatriz-dias/>. Acesso em: 19 mar. 2023.

DZIEKANIAK, G. V. Participação do bibliotecário na criação e planejamento de projetos de softwares: o envolvimento com a tecnologia da informação. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 2, n. 2, p. 37-56, 2004. DOI: 10.20396/rdbci.v2i1.2072 Acesso em: 23 abr. 2023.

FARINELLI, F. **Improving semantic interoperability in the obstetric and neonatal domain through an approach based on ontological realism**. 256p. Tese (Doutorado em Ciência da Informação) - UFMG, Belo Horizonte, 2017. Publicado em Inglês. Disponível em: [https://repositorio.ufmg.br/bitstream/1843/BUBD-AX2J5B/1/tese\\_ppggoc\\_fernandafarinelli.pdf](https://repositorio.ufmg.br/bitstream/1843/BUBD-AX2J5B/1/tese_ppggoc_fernandafarinelli.pdf). Acesso: 10 Mar. 2023.

FARINELLI, F. **Um diálogo entre o realismo ontológico e a engenharia de ontologias na construção de artefatos de representação**. In: ALMEIDA, M.B. *Representação do conhecimento, ontologias e linguagem: pesquisa aplicada em ciência da informação*. 1 ed. Curitiba, Brasil: Editora CRV, 2020. cap. 11, p. 277-294.

FARINELLI, F.; ELKIN, P. L. Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico. **Ciência da Informação**, Brasília, 46, n. 1, 2017.

FARINELLI, F.; ALMEIDA, M. B. Ontologias biomédicas: teoria e prática. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE - SBCAS 2019, 18º. 2019, Niterói, RJ, Brasil. **Anais .. 2019**, Niterói/RJ: Sociedade Brasileira de Computação. p. 93-140. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/29/97/247-1>. Acesso: 10 Mar. 2023.

FARINELLI, F.; SOUZA, A. D. D. Ontologias de alto nível: porque precisamos e como usar. **Fronteiras da Representação do Conhecimento**, Belo Horizonte, v. 1, n. 1, p. 174-202, 2021.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6ª ed. São Paulo: Atlas, 2008.

GILLILAND, A. J. Setting the stage: Building the metadata foundation. In: **Introduction to metadata** (pp. 1-20). Getty Publications, 2022.

HAWKINS, D. T.; LARSON, S.E.; CATON, B. Q. Information science abstracts: tracking the literature of information science. Part 2: a new taxonomy for information science. **Journal of the American Society for Information Science and Technology**, New York, v. 54, n. 8, p. 771-781, 2003.

HEATH, T.; BIZER, C. Linked data: evolving the web into a global data space. **Synthesis lectures on the semantic web: theory and technology**, v. 1, n. 1, p. 1-136, 2011.

HJØRLAND, B. Semantics and knowledge organization. **Annual review of information science and technology**, v.41, n. 1, p. 367-405, 2007.

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

KOBASHI, N. Y.; FRANCELIN, M. M. Conceitos, categorias e organização do conhecimento. **Informação & informação**, Londrina, 16, n. 2, p. 1-24, 2011.

LANCASTER, F. W. **Information retrieval systems**. 2<sup>a</sup> ed. New York: Wiley, 1978.

LANEY, D. 3D data management: Controlling data volume, velocity and variety. **META Group Research Note**, 6, p. 70-73, 2001.

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

LOUREIRO, M. F.; JANNUZZI, P. M. Profissional da informação: um conceito em construção. **Transinformação**, Campinas, v. 17, p. 1-30, 2005. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/217462>. Acesso em: 07 abr. 2023.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data**: a revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt, 2013. 242 p.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data**: Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013. 176 p.

MA, L.; XIAO, L. An overview of XML-based web development. **Journal of Computer Science and Technology**, v. 26, n.3, p. 351-362, 2011.

MCGUINNESS, D. L.; VAN HARMELEN, F. (ds.). OWL Web Ontology Language overview. **W3C recommendation**, v.10, n.10, 2004.

MELO, J. F.; ROCKEMBACH, M. Arquivologia e ciência da informação na era do Big Data: perspectivas de pesquisa e atuação profissional em arquivos digitais. **Prisma.com**, n. 39, p. 14-28, 2019.

MENDONÇA, F. M. **Ontoforinfoscience**: metodologia para construção de ontologias pelos cientistas da informação - Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). 2015. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, UFMG, Belo Horizonte, 2015. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUBD-A35H3K>. Acesso em: 01 Mai. 2016.

MENDONÇA, F. M.; SOARES, A. L. **Construindo ontologias com a metodologia ontoforinfoscience**: uma abordagem detalhada das atividades do desenvolvimento ontológico. *Ciência da Informação*, 46, n. 1, p. 43-59, 2017.

MENDONÇA, F. M.; ALMEIDA, M. B. D. **Ontoforinfoscience**: uma metodologia detalhada para construção de ontologias e sua aplicação no domínio da biomedicina. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB 2015), XVI, 2015, João Pessoa. **Anais..** João Pessoa: ANCIB, 2015.

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

MOUHA. Internet of Things (IoT). **Journal of Data Analysis and Information Processing**, v.9, p.77-101,2021 doi: 10.4236/jdaip.2021.92006.

PATEL-SCHNEIDER, P. F. . OWL web ontology language semantics and abstract syntax. **W3C recommendation**, v.10, n.10, 2005.

ODDONE, N.; GOMES, M.Y. F. S. de F. Uma nova taxonomia para a ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, Belo Horizonte, V, 2003. **Anais...** Disponível em: <http://hdl.handle.net/20.500.11959/brapci/172858>. Acesso em: 20 fev. 2023.

REIS, L. C. R.; DA FONSECA, M. I. Big Data: Um novo campo de atuação para bibliotecários. **Prisma.com**, n. 41, p. 231-250, 2020.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**, 1, n. 1, p. 96-105, 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/40838>. Acesso em: 15 maio 2023.

RODRIGUES, A. A.; NÓBREGA, E.; DIAS, G. A. Desafios da gestão de dados na era do big data: perspectivas profissionais. **Informação e Tecnologia (ITEC)**, Marília/João Pessoa, v. 4, n. 2, p. 63-79, 2017.

RUSCHER, L. Catálogo de produtos em marketplaces: relações com a Biblioteconomia. **Medium IA Biblio BR Grupo**. nov. 2022. Disponível em: <https://medium.com/ia-biblio-br/cat%C3%A1logo-de-produtos-em-marketplaces-rela%C3%A7%C3%B5es-com-a-biblioteconomia-bddf36e9ecef>. Acesso em: 19 mar. 2023.

SALES, L.F.; SAYÃO, L.F.; DA MOTTA, D.F. Modelagem de Relações Conceituais para a Área Nuclear. In: PROCEEDINGS OF V SEMINAR ON ONTOLOGY RESEARCH IN BRAZIL AND VII INTERNATIONAL WORKSHOP ON METAMODELS, ONTOLOGIES AND SEMANTIC TECHNOLOGIES (**ONTOBRAS-MOST**). **Proceedings...** 2012. Recife. p. 182-187.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em ciência da informação**, Belo Horizonte, v. 1, n. 1, 1996.

SAYÃO, L. F. Modelos teóricos em Ciência da Informação: abstração e método científico. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 82-91, jan./abr. 2001. E-ISSN 1518-8353. Disponível em: . Acesso em: 10 abr. 2023.

SEMELER, A.R.; PINTO, A.L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, v. 48, n. 1, 2019.

SILVA, A.R. **SEO semântico: fluxo de trabalho semântico**. 2d. Porto Alegre: Ed. do Autor, 2022. E-book.

SILVA, A.R. A criação de definições para o SEO Semântico. **Medium IA Biblio BR Grupo**. fev. 2023. Disponível em: <https://medium.com/ia-biblio-br/a-cria%C3%A7%C3%A3o-de-defini%C3%A7%C3%B5es-para-o-seo-sem%C3%A2ntico-a8b841d237d4>. Acesso em: 22 abr. 2023.

SILVA, S. D. M.; FARINELLI, F.; ALMEIDA, M. B., 2014, Belo Horizonte, Brasil. Um roteiro para modelagem conceitual de sistemas de informação baseada em princípios ontológicos. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: ALÉM DAS NUUVENS, EXPANDINDO AS FRONTEIRAS DA CIÊNCIA DA INFORMAÇÃO, XV, 2014. Anais..* Belo Horizonte. p. 4072-4095. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/189583>. Acesso em: 15 maio 2023.

SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da Informação em transformação: big data, nuvens, redes sociais e web semântica. **Ciência da Informação**, Brasília, v. 42, n. 2, 2013.

SOUZA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Modelos de representação semântica na era do Big Data. **Brazilian Journal of Information Science**, Marília, v. 12, n. 3, p. 34-40, 2018.

SOUZA, A. D. **Sistematização do processo de criação de definições formais em ontologias biomédicas: uma investigação no domínio das leucemias mieloides agudas**. 2015. 180 f. Dissertação (Mestrado em Ciência da Informação) Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2015. Disponível em : [ttp://hdl.handle.net/1843/BUBD-AAANE6D](http://hdl.handle.net/1843/BUBD-AAANE6D). Acesso em: maio 2023.

SOUZA, A.D; FARINELLI, F.; FELIPE, E. R. Formulação de definições utilizando o método Aristotélico. **Fronteiras da Representação do Conhecimento**, Belo Horizonte, v. 1, n. 1, p. 73–93, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/fronteiras-rc/article/view/35099>. Acesso em: 25 abr. 2023.

SOUZA, A. D.; ALMEIDA, M. B. Definitions in Biomedical Ontologies: the Method ONTODEF. **Journal of Biotechnology & BioResearch**, New York, v. 3, p. 1-2, 2022.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. D., 2010, English, Rome, Italy. **The KOS spectra: A tentative typology of knowledge organization systems**. Wurzburg: Ergon Verlag, 2010. p.122-128.

TARGINO, M,D,G. Quem é o profissional da informação?. **Transinformação**, Campinas, v.12, n.2, p.61-9, 2000. Disponível em:: <https://www.scielo.br/j/tinf/a/MhsxrLLcVF4jCBD7zWyxrbf/>. Acesso em: abr. 2023.

TAURION, C. **Big data**. Rio de Janeiro: Brasport, 2013.

TAYLOR, A. G. **The Organization of Information**. London: Libraries Unlimited. 2004. 417 p.

VALENTIM, M. L. P. Formação: competências e habilidades do profissional da informação. *In*: VALENTIM, M. L. **Formação do profissional da informação**. São Paulo: Polis, 2002. p.117-132.

VITORINO, E. V.; PIANTOLA, D. **Competência em informação**: conceito, contexto histórico e olhares para a Ciência da Informação. Florianópolis: Editora da UFSC, 2020. 205 p. Disponível em : <https://repositorio.ufsc.br/bitstream/handle/123456789/212553/E-book%20Compet%C3%Aancia%20em%20informa%C3%A7%C3%A3o%2031ago20.pdf?sequence=1&isAllowed=y>. Acesso em 08/04/2023.

W3C. **Semantic Web**. Disponível em: <https://www.w3.org/standards/semanticweb/>. Acesso em: 07 abr. 2023.