

ONTOLOGIA E WEB SEMÂNTICA: O ESPAÇO DA PESQUISA EM CIÊNCIA DA INFORMAÇÃO

Carlos Henrique Marcondes¹

marcon@vm.uff.br

Maria Luiza de Almeida Campos²

marialuizalmeida@gmail.com

RESUMO

Discute o atual projeto Web Semântica, em especial as ontologias, a partir de uma trajetória retrospectiva que remota à invenção do computador e da sua influência no imaginário da sociedade contemporânea como uma máquina dotada de inteligência. As tendências, limitações e desafios atuais do projeto e dessa visão, segundo diferentes autores, são apresentados e discutidos. São identificadas possíveis áreas de atuação para a Ciência da Informação, em especial para a construção de ontologias, a partir sua trajetória de pesquisa em áreas como organização/modelização de domínios de conhecimento e processamento semântico de informações por computadores.

Palavras-chave: Web semântica; Ontologia; Pesquisa em Ciência da Informação; Modelização de domínios de conhecimento; Processamento semântico de informações.

ONTOLOGY AND SEMANTIC WEB: THE FIELD OF THE RESEARCH IN INFORMATION SCIENCE

ABSTRACT

Discusses the Semantic Web project, in special ontology, from a viewpoint which traces back to the invention of computer and its influence in social imaginary as an intelligent machine. Project's actual trends, limits and challenges of this vision, according to different authors, are presented and discussed. Possible areas where Information Science could act, mainly in ontology development, are identified, from its research experience in topics as knowledge organization/modeling and semantic processing of information by computers.

Keywords: Semantic Web; Ontology; Research in Information Science; Domain knowledge modeling, semantic processing of information by computers

¹ Depto. de Ciência da Informação/UFF

² Depto. de Ciência da Informação/UFF

1. DAS INSTITUIÇÕES DE GUARDA E DISSEMINAÇÃO DO CONHECIMENTO ÀS ONTOLOGIAS

El universo (que otros llaman la Biblioteca) se compone de un numero indefinido, y tal vez infinito, de galerias hexagonales, com vastos pozos de ventilación em el médio, cercados por barandas bajísimas (BORGES, 2007, p. 86).

A sociedade atual tem no conhecimento um de seus pilares econômicos e simbólicos. No amplo espectro da Cultura, o conhecimento científico, desde a Modernidade, vem evoluindo segundo uma dinâmica própria e passando desde então a se imbricar cada vez mais com o sistema produtivo (GÓMEZ, 1987). Hoje a sociedade não produz sem o aporte do conhecimento em geral e, especificamente, do conhecimento científico. Não é por outra razão que a sociedade atual é chamada de sociedade da informação, sociedade do conhecimento ou “modo informacional de desenvolvimento”, conforme Castells (1999, p.54). Meios econômicos, sociais, políticos e tecnológicos são aportados pela sociedade atual para gerir seu acervo de conhecimento. O computador e a Internet são, muito justamente, tanto o símbolo quanto a principal ferramenta da sociedade da informação.

A sociedade humana inventou a escrita, no Egito e na Mesopotâmia, por volta de 3000 a.C. A escrita e, mais especificamente, os documentos, são artefatos sociotécnicos que permitem a permanência e a apropriação social de fatos na ausência destes, perpetuando-os através do tempo e (por seu caráter portátil) do espaço. Dada a sua materialidade, também podem ser acumulados, permitindo sua re-apropriação social posterior. Desde então que o sonho ou o delírio de acumular, controlar, inventariar de forma total e organizar seu legado de memória coletiva têm sido uma obsessão da humanidade. Este sonho se manifesta desde a biblioteca de Nínive, na Babilônia, passando pela biblioteca de Alexandria, pelos escritórios dos monges copistas medievais, pela *Encyclopédie* de Diderot e D'Alembert, pela bibliografia universal de Otlet. Esse sonho é talvez o que hoje esteja sendo re-colocado com a visão da Web Semântica (BERNERS-LEE, 2001).

A reprodução social do conhecimento através da guarda, disseminação e uso é uma questão com implicações tão profundas para a sociedade humana, a ponto dela, ao longo da sua evolução, ter dedicado grandes esforços e criado instituições especiais para manutenção e expansão do Conhecimento e da Cultura. Além do sistema de ensino e pesquisa, Museus, Bibliotecas e Arquivos são algumas dessas instituições, criadas com

a missão precípua de preservar e disseminar a Cultura. As atividades dessas instituições e as metodologias aí utilizadas compõem o campo prático da Ciência da Informação.

A Internet adquire sua face atual com o surgimento do hipertexto e da Web, a “teia global” criados por Tim Berners-Lee no CERN em 1989. A Web transforma a Internet num gigantesco sistema de informações à escala mundial. No entanto o crescimento gigantesco da Web a partir daí colocou novos e inéditos problemas para o acesso a informação aí disponibilizada. É muito fácil para qualquer um publicar na Internet, o que fez com que a rede tivesse um crescimento desordenado e caótico e que encontrar a informação adequada passasse a ser o principal problema cultural, econômico e científico da atualidade. Nunca a humanidade dispôs de tanta informação e ao mesmo tempo nunca foi tão difícil e problemático encontrar a informação relevante. Esse é o principal problema da atual economia da Informação.

Ao endereçar este problema, Tim Berners-Lee propõe a visão da Web Semântica, uma extensão da Web atual, formada por documentos compreensíveis unicamente por pessoas, para uma Web em que documentos seriam auto-descritíveis, de forma que seu conteúdo possa ser “compreendido” por programas, os agentes de “software”*, que assim poderiam “raciocinar” e fazer “inferências” sobre o conteúdo de documentos, ajudando as pessoas em diferentes tarefas de recuperação de informações que exijam, raciocínio, decisões, inferência de conclusões a partir de informações não explicitamente disponíveis ou de informações contextuais. Segundo Tim Berners-Lee (2001, p. 2): “A Web Semântica irá trazer estrutura ao conteúdo das páginas Web, criando um ambiente onde agentes de “software” navegando de página em página poderão desenvolver tarefas sofisticadas para os usuários”.**

O conhecimento científico, tão importante para nossa sociedade, vinha sendo guardado, preservado e disponibilizado desde a antiguidade nas coleções armazenadas em bibliotecas. O estatuto da guarda, preservação e disponibilização da cultura, dos conhecimentos em geral, em especial, do conhecimento científico, tão caro para as bibliotecas enquanto instituições vêm passando por um desafio com o surgimento da Web.

* http://en.wikipedia.org/wiki/Software_agent

** Tradução nossa.

Artigos científicos são o veículo através do qual são disseminados os novos conhecimentos. Desde a publicação das *Philosophical Transactions* da Royal Society na Inglaterra do século XVII, coleções de artigos científicos armazenadas nas bibliotecas eram os repositório dos novos conhecimentos. As bibliotecas, e hoje as bibliotecas digitais vêm tendo este papel, de repositórios do conhecimento da humanidade. No entanto hoje as publicações científicas na Web, apesar do avanço das TIs, são ainda calcadas no modelo impresso.

Artigos científicos são bases de conhecimento, mas esse conhecimento é textual e só pode ser processado por seres humanos. Existem dois obstáculos para o acesso e utilização em larga escala deste conhecimento: o grande número de publicações, a chamada “explosão informacional” (MARCONDES, 2001), que atinge mais alto grau com o surgimento da Web e das publicações eletrônicas; e o fato desse conhecimento estar inserido no texto dos artigos de forma não estruturada, legível somente por pessoas. São estas que têm que comparar, avaliar a coerência, relacionar, inferir e citar o conhecimento contido em textos. O formato textual não estruturado impede que este conhecimento seja processado por programas, como é a proposta da Web Semântica.

Hoje grupos de pesquisadores lançam-se na tarefa de sistematizar e estruturar o conhecimento científico em domínios específicos e disponibilizá-lo publicamente na Web, através das chamadas ontologias, de modo a permitir que comunidades científicas compartilhem informações sobre domínios específicos. Nas Ciências da Saúde, o consórcio OBO - Open Biomedical Ontology - (<http://obo.sourceforge.net/>), congrega mais de 70 diferentes ontologias. Ontologia é hoje uma área de pesquisa emergente.

Como apresentado em diversos estudos, ontologia* (CORAZZON, 2000; GRUBER, 1993; GUARINO, 1997, 1998 a; SWARTOUT; TATE, 1999; VICKERY, 1997; SMITH, 2002) como instrumento de representação de conhecimento, surge no âmbito da Inteligência artificial na década de 90. Para os sistemas de Inteligência Artificial, o que existe é o que pode ser representado. Quando o conhecimento de um domínio é representado em uma linguagem declarativa, o conjunto de objetos que podem ser representados é chamado de universo do discurso. Foi nesse sentido que surgiram as ontologias, com o intuito de descrever dados manipulados por programas,

* O termo Ontologia tem suas origens na Filosofia, como o estudo do ser, de suas características e de suas condições de existência

através da definição de um conjunto de termos que pudessem representar domínios e tarefas a serem executadas por estes programas.

Atualmente as ontologias vêm sendo utilizadas em diversas áreas. Guarino (1998 a) em seu trabalho afirma que esta ontologia com “o” minúsculo, diferentemente da ontologia com “O” maiúsculo, campo de estudos da Filosofia (nota), tem sido reconhecida em diversos campos de pesquisa, tais como: engenharia do conhecimento, análise orientada a objeto, recuperação e extração de informação, organização e gerência de conhecimento, projeto de sistemas baseados em agentes, projeto de banco de dados, entre outros. Barry Smith (2002), discutindo a relação entre Ciência e Ontologia (enquanto domínio de conhecimento preocupado com a natureza dos seres), afirma que enquanto o papel da ciência é “explicar” a natureza, o papel da Ontologia seria vir a seguir para descrever, organizar e sistematizar o conhecimento obtido pelas descobertas científicas. Este parece ser um lugar a ser ocupado também pela Ciência da Informação, vinda de uma longa tradição de organização de domínios de conhecimento, aplicada originalmente à organização de repertórios documentais.

Definições com origem na Ciência da Informação, mais ligadas ao foco do uso das ontologias na Web Semântica, são as seguintes:

Ontologia é definida como uma especificação formal e explícita de um conceitualização compartilhada. Fornece uma compreensão comum e compartilhada de um domínio que pode ser comunicada a pessoas e sistemas* (DING, 2002b, p. 375).

Uma conceitualização parcial de um domínio de conhecimento, compartilhada por uma comunidade de usuários, definida em linguagem formal, processável por máquina, para o objetivo explícito de compartilhar informação semântica entre sistemas automatizados.* (JACOB, 2003, p. 20).

Nestas definições se destacam as noções de “conceitualização”, “compartilhada”, num determinado “domínio” e de representação “formal”, “processável por programas”.

Uma conceitualização é uma abstração, uma visão simplificada do mundo que se representa para satisfazer um ou mais dos seguintes propósitos: "permitir que múltiplos agentes compartilhem seu conhecimento; ajudar as pessoas a compreender melhor certa área de conhecimento; ajudar pessoas a atingir um consenso no seu entendimento

* Tradução nossa

sobre uma área de conhecimento” (SMITH, apud FALBO, 1998). Em Lógica, uma conceituação identifica o objeto e relações que existem no universo lógico. (WEINSTEIN, 1998)

A literatura, muitas vezes, vem denominando ontologias pequenas estruturas de conceitos. Essas estruturas, apesar de possuírem conceitos e relações, não possuem definição na forma de axiomas dos seus conceitos e, na maioria das vezes, não são árvores, mas grafos. Estes tipos de ontologias são chamadas, também, de ontologias informais (WEINSTEIN, 1998) ou ontologias lingüísticas (GUARINO, 1998 b). Na verdade, o termo ontologia vem sendo usado para caracterizar qualquer classificação de um domínio ou uma lista de termos definidos.

Em Inteligência Artificial o que se denomina por ontologia é considerado ontologia formal, ou seja, aquela que define vocabulário com o uso da Lógica. A sintaxe exata e semântica dependem da linguagem de representação, que é expressa sintaticamente com uma linguagem equivalente à lógica de primeira ordem. Desta forma, uma ontologia consiste em termos, definições, e axiomas relativos a eles. (GRUBER, 1993)

A diferença entre uma ontologia lingüística e uma formal é que as ontologias lingüísticas constroem redes semânticas entre palavras, onde estão em jogo cadeias de associações que, na maioria dos casos, não estão baseadas em relações lógicas. Por exemplo, em ontologias lingüísticas pode-se considerar que couro sintético é um tipo de couro. Mas se levarmos em conta a definição de couro, ou seja, seu conteúdo conceitual, couro é um produto de origem animal, não se pode estabelecer um tipo de subordinação entre couro e couro sintético que não esteja baseada em uma relação lógica. Uma ontologia formal por sua vez elabora redes conceituais com relações que formam cadeias lógicas e ônticas como, por exemplo, as relações de generalização e agregação respectivamente.(CAMPOS et al, 2005).

Ontologias foram pensadas no contexto da Web Semântica para trabalharem associadas a sistemas de informação na Web - os “web services” (BREITMAN, 2005, p.141) - registrando a semântica do funcionamento dos mesmos de modo genérico, usando linguagens padronizadas como a OWL - Ontology Web Language, um padrão do W3C -, de modo a tornar sua semântica de operação “compreensível” por programas agentes de “software”, permitindo assim sua interoperabilidade com outros sistemas distintos.

Um exemplo dessas interações entre sistemas diferentes ou “web services”, cuja semântica estivesse especificada em ontologias específicas, seria um sistema bibliográfico que respondesse ao seguinte exemplo consulta complexa:

- Que artigos, publicados em periódicos em Ciências da Saúde com avaliação por pares, que tenham fator de impacto maior que X, são teóricos, tem como autores pesquisadores cuja afiliação seja numa das 100 primeiras universidades mundiais, de acordo com o “raking” de Y?

Um sistema/agente capaz de responder a esta consulta teria que interagir com ontologias associadas a pelo menos três dos seguintes outros sistemas/“web services”:

- a- sistemas de registro e “ranqueamento” de universidades e instituições de pesquisa, que forneceriam meios de identificar das primeiras 100 de instituições de acordo com o “ranking” Y;
- b- sistemas de publicação de periódicos eletrônicos, que forneceriam meios para identificar periódicos em Ciências da Saúde com fator de impacto maior que Y;
- c- finalmente os parâmetros fornecidos pelos dois sistemas anteriores forneceriam meios para formular consultas num sistema de gestão da biblioteca digital ou base de dados onde os artigos estariam armazenados.

A Ciência da Informação, ciência dos conteúdos registrados, das suas transferências com vistas à sua apropriação social, vem de uma longa tradição teórica, metodológica e prática que converge para as questões atuais colocadas pela proposta da Web Semântica e para a construção de ontologias. Áreas de pesquisa recorrentes na Ciência da Informação são a organização e modelização de domínios de conhecimento (CAMPOS, 2004), aplicada originalmente para viabilizar o acesso e a organização de repertórios documentais e processamento semântico de informações em computadores. Verifica-se que ao falar em ontologias estas questões são altamente interligadas. Contribuições significativas da CI, já históricas, nestas áreas são as de Shera, (1957), Luhn (1960), Outlet (1989), Ranganathan (1967), Dahlberg (1978) e Hjørland (2002).

O desenvolvimento cada vez mais generalizado da Web Semântica das ontologias em diferentes domínios do conhecimento e os problemas aí colocados demandam o resgate de todos estes aportes metodológicos da Ciência da Informação. Essa relação da CI com as ontologias é cada vez mais percebida pelos pesquisadores da área (SOERGEL, 2000), (THE SEMANTIC WEB, 2003), (DING, 2002a, 2002b). Este artigo se propõe a pontuar alguns dos problemas e desafios hoje enfrentados pelo projeto Web Semântica e possíveis contribuições da Ciência da Informação. Esta

estruturado assim: nas seção seguinte, os problemas do atual projeto da Web Semântica são discutidos. Na seção 3 possíveis interseções entre as questões apontadas e questões levantadas em toda a trajetória da CI são identificados, especialmente em aspectos como processamento semântico de informações em computadores e a organização e modelagem de domínios de conhecimento. Por fim, da Seção 4 são apresentadas conclusões.

2. Web Semântica e o registro de conhecimentos em ontologias: problemas e desafios

Nenhum tipo de conhecimento, mesmo que pareça-nos tão natural, por exemplo, quanto a teoria, é independente do uso de tecnologias intelectuais. (LÉVY, 1993, p. 75).

O processamento semântico de informações por computadores esbarra hoje, no contexto da Web Semântica e das ontologias, em alguns obstáculos que remontam às origens da computação moderna. Seria o computador, chamado nos primórdios da informática, de “cérebro eletrônico”, capaz de “pensar”? São viáveis as propostas de disciplinas como a Inteligência Artificial ou a do atual projeto da Web Semântica, que coloca explicitamente: “A Web Semântica não é uma Web separada, mas sim uma extensão da Web atual, na qual à informação é dado um significado definido, favorecendo que computadores e pessoas trabalhem em cooperação”^{*} (BERNERS-LEE, 2, 2001). O projeto da Web Semântica tem objetivos ambiciosos: “A Web Semântica não é meramente uma ferramenta para conduzir tarefas individuais que nós vimos discutindo até agora. Mais que isso, se propriamente planejada, a Web Semântica poderá apoiar a evolução do conhecimento humano como um todo”^{*} (BERNERS-LEE, 2, 2001).

Questões como estas têm fascinado o imaginário da humanidade na atualidade, alimentado por um grande número de obras de ficção. O computador eletrônico é filho direto das grandes calculadoras mecânicas empregadas desde o século XIX para o cálculo de artilharia dos navios de guerra (BRETON, 1991). Nas décadas de 20, 30 e 40 do século passado, à tecnologia de cálculo veio se somar à tecnologia eletrônica. O modelo do computador, concebido anteriormente à construção do primeiro computador

^{*} Tradução dos autores

real ainda na década de 30, é a chamada máquina de Turing, devida a este matemático. É uma máquina capaz de realizar um procedimento sistemático de cálculo, um algoritmo. Nos primórdios da computação esta se caracterizava por uma forte ênfase no processamento, cálculos complexos sobre poucos dados.

O computador logo extrapola os usos iniciais previstos por seus criadores, de ser uma máquina que realiza cálculos complexos com rapidez e precisão. No entanto o paradigma do cálculo, do processamento, do algoritmo, marcante na construção dos primeiros computadores, permanece até hoje na sua estrutura de funcionamento e, se foi uma das razões de seu sucesso, é também hoje uma das suas maiores limitações. Várias camadas de “software” - interpretadores, linguagens de alto nível, sistemas operacionais, interfaces gráficas - vêm se interpondo entre a “linguagem de máquina”, única compreensível e executável pelos processadores, e os usuários durante a evolução do computador. No entanto, desde o ENIAC* até os modernos computadores atuais, computadores são no fundo máquinas que trabalham dentro de um rigoroso e sistemático processamento repetitivo, passo a passo, algorítmico (MAGALHÃES, 1997). Estas limitações são percebidas principalmente quando se trata de realizar processamento semântico de conteúdos.

Os pesquisadores dos primórdios da invenção dos computadores, encantados com as possibilidades da nova máquina em realizar cálculos e operações lógicas, chegaram a afirmar ser ela funcionalmente equivalente ao cérebro humano e que o computador seria capaz de “pensar” (DUPUY, 1996). É interessante ver a polêmica sobre esta questão envolvendo Hubert Dreyfus e Edward Feigenbaun, dois pioneiros da Inteligência Artificial (PESSIS-PASTERNAK, 1993). O desenvolvimento da idéia de que pensar é calcular, de que o computador poderia simular o funcionamento do cérebro humano levou, ainda na década de 50, ao surgimento da primeira Cibernética e da Inteligência Artificial. Esta especificamente, surge tendo como projeto simular em computadores procedimentos ditos inteligentes, como raciocínio, decisão, resolução de problemas, compreensão de linguagem natural, etc. Para a Inteligência Artificial voltar-se para a Lógica enquanto método de simular processos inteligentes foi uma opção óbvia. Durante a década de 70 e 80 a Inteligência Artificial apoiou-se fortemente numa

* Electronic Numeric Integrator and Calculator, o primeiro computador eletrônico, que começou a funcionar em 1944.

longa tradição do pensamento ocidental que remota aos gregos (TENÓRIO, 2001, p. 30), do uso da lógica como método de raciocínio que permitiria conduzir à verdade ou, na prática, resolver um problema.

A Inteligência Artificial baseou-se fortemente na chamada programação em lógica como método de resolver problemas, que consiste basicamente num procedimento algorítmico: “Essa proposta é colocada por Hilbert da seguinte forma: a descoberta de um método para estabelecer a verdade ou falsidade de qualquer sentença na linguagem da lógica formal chamada cálculo de predicado” (TENÓRIO, 2001, p. 40). A proposta do matemático Hilbert é concretizada no chamado “Princípio da Resolução”, devido ao também matemático Robinson, na década de 60, que consiste exatamente num procedimento algorítmico para provar a validade ou falsidade de um conjunto de sentenças expressas em lógica (GENARO, 1985, p. 75). Este conjunto de “sentenças expressas em lógica” viria a se constituir na base de conhecimento dos sistemas especialistas.

Após a criação de programas experimentais na década de 60, as realizações práticas da IA foram na construção de sistemas especialistas, ou sistemas peritos, que objetivavam simular em computadores as habilidades de um especialista na resolução de problemas específicos. Os sistemas especialistas tinham como componentes básicos uma base de conhecimento e um motor de inferência. O Princípio da Resolução mencionado anteriormente é a base dos motores de inferência dos sistemas especialistas. A base de conhecimento continha, de forma codificada, o conhecimento obtido de um especialista, expresso como um conjunto de sentenças expressas em lógica.

A obtenção e codificação do conhecimento de um especialista humano - a chamada “aquisição do conhecimento” - era um primeiro problema na construção e na eficiência dos sistemas especialistas e continua sendo hoje com as ontologias. Forsythe (1989) define aquisição de conhecimento como:

Nesse “paper” nós endereçamos o problema de identificar e mitigar dificuldades de comunicação no processo obter informação de um ou mais especialistas humanos para o propósito de construir uma base de conhecimento. A aquisição de conhecimento veio a se mostrar o gargalo no processo de construção de sistemas especialistas.

A tarefa algumas vezes chamada de “extração do conhecimento de especialistas humanos”, pode ser lenta, ineficiente e frustrante para especialistas e engenheiros de conhecimento.

Sobre a base de conhecimento atuava o motor de inferência, simulando o raciocínio do especialista ao realizar inferências sobre o conhecimento contido na base de conhecimento. Buchanan (1988, p. 32) no entanto, destaca limitações e especificidades do escopo de aplicações dos sistemas especialistas: “Sistemas especialistas contam mais com formulações de casos especiais que em princípios básicos e gerais”^{*}.

As bases de conhecimentos dos sistemas especialistas podem ser consideradas as antecessoras das modernas ontologias. Ambos os componentes dos sistemas especialistas - base de conhecimento e motor de inferência - eram altamente integrados, em detrimento de um possível re-uso da base de conhecimento para outras aplicações.

Pouco a pouco, a partir da década de 90, a ênfase nos aspectos lógicos da resolução de problemas cede lugar ao desenvolvimento de bases de conhecimento - ou ontologias -, consistentes e completas sobre domínios da realidade, independentes de aplicações específicas. Essa mudança é comentada por Guarino (1995):

Pesquisadores em IA parecem estar muito mais interessados na natureza do raciocínio que na natureza do mundo real. Recentemente o valor potencial de bases de conhecimento (ou ontologias) independentes de tarefas passíveis de serem integráveis, começa a ser reconhecido de diversas maneiras^{*}.

A proposta da Web Semântica prevê as ontologias como um dos pilares da sua arquitetura. As ontologias enquanto bases de conhecimento sobre domínios específicos estariam associadas a sistemas ou “web services”, guardando o conhecimento sobre e a semântica desses sistemas, de modo que programas inteligentes, os “agentes de software” pudessem interagir sem restrições com esses sistemas, tornando-os interoperáveis.

Veltman (2004) afirma existir uma linha evolutiva direta entre o projeto da IA e o da Web Semântica. Aponta para as contradições desse projeto e de suas limitações no que diz respeito a lidar com problemas relativos à cultura e a arte, campos onde são

^{*} Tradução dos autores

valorizados os múltiplos significados e interpretações, condicionados social e culturalmente.

Podem programas compreender significados? Veltman (2004), ao discutir criticamente o projeto da Web Semântica, questiona esta possibilidade:

Hoje a web semântica lida com significados de uma maneira muito restrita e oferece soluções estáticas. Isso é adequado para muitos propósitos nas transações em ciência, tecnologia e negócios que requerem comunicação máquina a máquina, mas não responde às necessidades da cultura. *

A viabilidade deste tipo de aplicação depende de que esse conhecimento, registrado em ontologias distintas, seja coerente entre si, não contraditório, construído sobre bases teóricas e metodológicas sólidas. No âmbito dos estudos em ontologias na perspectiva da Web Semântica são as ontologias formais que assumem este papel. As ontologias formais ligadas ao conceito de modelagem de conhecimento podem ser consideradas um mecanismo de representação, como um meio próprio de observação do conhecimento de um dado domínio. O objetivo da modelagem de conhecimento é elaborar uma conceituação da porção do mundo em estudo.

Toda ontologia formal está pautada em uma taxionomia, como uma "espinha dorsal" de qualquer domínio de conhecimento. (GUARINO, 1998 a).

Na prática, a Ontologia Formal pode ser entendida como a teoria das distinções a priori sobre: as entidades do mundo (objetos físicos, eventos, regiões, quantidades de matéria); as categorias de meta-nível para modelar o mundo (conceitos, propriedades, qualidades, estados, papéis e partes). (GUARINO, 1998a).

Smith chama atenção para algumas destas questões, no domínio específico e altamente desenvolvido das ontologias biomédicas:

Princípios formais que governam boas práticas em classificação e definição foram por muito tempo negligenciados na construção de ontologias biomédicas, resultando em conseqüências negativas importantes para a integração de dados e alinhamento de ontologias. * (SMITH, 2004, p. 1).

Uma série de "papers" recentes mostrou que a Gene Ontology (GO), o mais conhecido recurso taxonômico nesse campo, é marcada por equívocos em certos tipos característicos, que são resultado de uma falha em endereçar princípios ontológicos básicos * (SMITH, 2006, p. 1).

* Tradução dos autores

O problema é que quando a OBO e ontologias similares incorporam tais relações elas tipicamente o fazem de maneira informal, de maneira que as interconexões lógicas entre as várias relações empregadas não são claras e mesmo as relações `é_um_e_parte_de` não são sempre usadas de uma maneira consistente tanto dentro de uma mesma quanto entre diferentes ontologias.* (SMITH, 2005, p. 2).

Uma das bases da construção de qualquer ontologia é a sua estrutura taxonômica. Guarino (2000) ressalta deficiências neste aspecto: “Talvez o problema mais comum que nós temos visto na prática com ontologias seja que, embora se espere que elas dêem ordem e estrutura à informação, sua estrutura taxonômica é frequentemente pobre e confusa”*.

Questões metodológicas como as colocadas por Barry Smith e Nicola Guarino comprometem o pressuposto básico da Web Semântica e do uso de ontologias: o processamento semântico desse conhecimento por computadores e reduzindo o uso de ontologias hoje ao de meras bases terminológicas para indexação manual, como são atualmente a UMLS para dados bibliográficos e a Gene Ontology para bases de dados computacionais relativas ao sequenciamento genético.

Às inconsistências semânticas e conceituais na construção de ontologias se somam as limitações do processamento semântico por computadores. Os procedimentos do assim chamado alinhamento de ontologias por computadores (DE BRUIJN, 2006), são na verdade processamento sintático de reconhecimento de padrões - identificação de “strings”, termos idênticos ou quase idênticos - em ontologias diferentes. Todas estas deficiências tornam as ontologias atuais empreendimentos isolados, com pouca ou nenhuma interoperabilidade, comprometendo assim a compatibilização de ontologias distintas e sua integração, o que potencializaria sua utilização. Os resultados do uso de ontologias para processamento por “agentes de software” têm sido ainda pouco significativos.

Outra questão fundamental a ser endereçada, na medida em que o conhecimento científico em áreas de ponta começa a ser registrado de forma crescente em ontologias legíveis por computadores e mantidas por comunidades restritas, é a validação de conhecimento. Muitas ontologias são “curadas”, ou seja, tem um curador, um especialista que valida seu conteúdo. Mas esses mecanismos são ainda informais e precários. Pode-se olhar para retrospectivamente para toda a experiência acumulada na validação do conhecimento científico quando o principal veículo para veiculá-lo eram os periódicos em papel. As críticas aos mecanismos de “peer-review” são contundentes

(HOROBIN, 1999), (CAMPANÁRIO, 1998), (CAMPANÁRIO, 2002) e devem ser re-colocadas quando ao se criarem mecanismos de avaliação do conhecimento registrado em ontologias.

3. Ontologias e o saber da Ciência da Informação

O desenvolvimento de ontologias demanda uma convergência de saberes, uma verdadeira interdisciplinaridade, na qual a Ciência da Informação pode vir a ter uma contribuição importante.

Tradicionalmente, a Ciência da Informação nasce já preocupada com o processamento semântico de informações por computadores. Saracevic (1995), discutindo a constituição da Ciência da Informação, afirma ser ela uma disciplina indissociavelmente ligada às tecnologias da informação. As décadas de 50 e 60 marcam as primeiras tentativas de usar o computador para processar conteúdos; são dessa época as experiências de pioneiros da Ciência da Informação como as de Shera (1957) com relação aos “semantic codes”, e Luhn (1960), no sentido de viabilizar o processamento semântico de informações por computador.

Uma de suas bases é a especificação semântica de relações. Relações estão na base do potencial dos computadores realizarem inferências. Esse termo deve ser entendido aqui não de forma literal, como uma atividade humana altamente sofisticada e criativa, mas simplesmente a capacidade de um programa obter novas informações a partir de informações dadas, ou seja, relacionadas, por ex, dado A, espera-se sempre B. A questão das relações é também explorada pela Ciência da Informação há tempos. A proposta de indexação relacional de Farradane (1980) valoriza especialmente as relações: “Significado, considerado como relações entre termos ...”*. Brookes (1980) é outro autor da CI que ressalta a importância das relações: “o conhecimento consiste numa estrutura de conceitos ligados por relações e a informação é uma pequena parte dessa estrutura”*. A pesquisa na semântica das relações tem aplicações na indexação. A proposta de relacionar termos de indexação para melhorar a recuperação de informações é descrita em Kajikawa (2006). Trabalhos como os de Perrault (1994) e a recente revisão de Khoo (2007) sobre relações apresentada na ARIST, mostram o quanto esse

* Tradução dos autores

tema é relevante na Ciência da Informação e vêm sendo explorado sistematicamente. Em Marcondes (2006) as relações são a base do conhecimento contido em artigos científicos. Inventariar e explorar a semântica de relações específicas em diferentes domínios como vêm sendo feito pela CI é essencial para viabilizar o processamento de conteúdos por computadores. Trabalhos relativos ao estabelecimento de classificações entre relações em ontologias e nas atividades tesaurográficas tem sido desenvolvido também no escopo da Ciência da Informação em nível nacional. (SALES, CAMPOS, GOMES, 2006). Além disso, as relações são de fundamental importância para as atividades de reuso em ontologias (Campos, 2007).

A estes saberes se somam os relativos à modelização de domínios específicos, onde a Ciência da Informação traz contribuições significativas como “sense making”, de Dervin (1983), de Hjörland (2002), relativos a metodologias que poderiam ser usadas para aquisição de conhecimentos, na proposta de categorias que ajudem a estruturar de forma consistente esses domínios, como em Ranganathan e Dahlberg (1978a), (1978b).

3.1. A Modelização de domínios de Conhecimento e sua importância para a elaboração de ontologias

Recentemente, ontologias passaram a receber atenção especial por parte da comunidade de Ciência da Informação. Uma ontologia define um vocabulário comum para uma comunidade que precisa compartilhar informação em um determinado domínio. Inclui definições de conceitos básicos no domínio e as relações entre estes de forma que sejam interpretáveis por máquina.

No entanto, apesar de propostas de linguagens e ferramentas de representação e construção de ontologias no domínio Ciência da Computação, estas propostas, em geral, ainda não contemplam diretrizes satisfatórias de identificação dos conceitos e relacionamentos entre estes, de criação de definições consistentes, nem tampouco de modelização de domínios associados. Por conseqüência, as ferramentas têm pouco a contribuir no sentido de orientação do usuário no processo de construção da ontologia, assim como em diretrizes para a construção de ontologias de qualidade. (FERNANDEZ, et al, 1997; DING, 2002a, 2002b; FERNÁNDEZ-LÓPES, 2001; GANGEMI, 1999 ; GRÜNINGER; FOX, 1995; GUARINO, 2002 ; METAXIOTIS, 2001; SURE et al, 2002; PASLARU-BONTAS, 2007; USCHOLD ; KING, 1995)

Algumas das tradicionais habilidades do profissional da informação - construção de tesouros e vocabulários controlados, concepção de classificações/taxionomia, organização da informação - contribuem para a próxima etapa de desenvolvimento Web semântica, onde as ontologias possuem papel de destaque. Estas habilidades estão apoiadas em bases teóricas e metodológicas no âmbito da Ciência da Informação. Ao longo de sua trajetória, a Ciência da Informação desenvolveu um arcabouço teórico e metodológico consubstanciado em obras como as de Otlet (1989), Ranganathan (1967), Dahlberg (1978a), Hjørland (2002) e do CRG - Classification Research Group (WILSON, 1972), entre outros, que potencialmente têm grande aplicabilidade no desenvolvimento de ontologias. (CAMPOS, 2005).

Neste estudo vamos abordar as questões relacionadas à modelização de domínios, que no âmbito das ontologias, são representadas por taxonomias, que formam a sua espinha dorsal, como citado por Guarino - 'Toda ontologia formal está pautada em uma taxionomia, como uma "espinha dorsal" de qualquer domínio de conhecimento'. (GUARINO, 1998 a).

Em ontologias, diferentemente de outras linguagens, também denominados de vocabulários controlados, que também tem por função permitir a representação de informação em bases compartilhadas, como tesouros, por exemplo, é necessário que se possa produzir o raciocínio inteligente através de mecanismos de inferência que dão suporte à manipulação do conhecimento explicitado. Desta maneira, os elementos que formam as ontologias, conceitos e seus relacionamentos devem ser agregados de forma consistente formando um sistema de conceitos, onde as partes estão relacionadas de tal forma, que são vista de forma sistêmica.

Entretanto, é importante acrescentar que uma ontologia é mais que um vocabulário controlado ou ainda que um tesouro. Conforme explicitaram Campos e colegas (2005):

Muita confusão se tem feito em torno do conceito de ontologia, que não pode ser considerado somente como um vocabulário controlado. Uma ontologia possui informações de natureza distinta, ou seja: terminológica - possui um conjunto básico de conceitos e relações-; e assertivas aplicadas aos conceitos e relações, que constitui um conjunto de axiomas, diferentemente de um tesouro." (CAMPOS, M. L. M; CAMPOS, M. L. A.; CAMPOS, L. M., 2005)

Se considerarmos uma ontologia como um sistema de conceitos interrelacionados, podemos dizer que estas contém basicamente três componentes, quais

sejam: 1. Um conjunto básico de conceitos e relações; 2. Uma forma de representação; 3. Um conjunto de assertivas.

No que concerne ao primeiro componente, verificamos que é neste sentido que vem sendo discutida a semelhança entre tesauros e ontologias. Tesauros são linguagens documentárias que possuem também termos e relações. Entretanto, diferentemente das ontologias, as relações em tesauros são menos especificadas. Mesmo os tesauros de vertente conceitual que avançam sobre os com base somente alfabética, não especificam suas relações de modo a possibilitar o raciocínio automatizado (CAMPOS, 2001), como é um dos requisitos para que uma ontologia venha a ser definida como formal.

Os tesauros conceituais de vertente européia apresentam relacionamentos de natureza lógica (TG/TE), de natureza ôntica (TGP/TEP; TA/TA; TO/TO) e de equivalência (USE/UP); além disso, explicitam para cada termo uma definição que é legível somente por humanos.

Entretanto, as ontologias necessitam de relações mais explicitadas para possibilitar inferências. Se para os tesauros é importante saber se entre dois termos existe uma relação de parte/todo. Nas ontologias devemos identificar que tipo de parte e de todo esses termos representam, por exemplo: se é um objeto integral/componente, como: roda é uma parte do carro; se é um membro/coleção, como : uma árvore é parte de uma floresta, mas mantém uma identidade própria; se é massa/ porção, como: o pedaço de uma torta. Por outro lado as relações associativas (relações funcionais) que no tesouro somente são identificadas pelo simbolismo (TA/TA), e são consideradas relações diáticas (SALES, CAMPOS, GOMES 2006) nas ontologias elas são mais explicitadas e expressam o tipo de relacionamento que existem entre os dois termos, ou seja, não adianta, como nos tesauros, somente indicar que no âmbito da Ciência da Saúde existe uma relação associativa entre um “Barbeiro” e a “Doença de Chagas”, ou seja entre uma entidade e um processo, mas que o relacionamento que se coloca entre esses dois elementos é aquele onde uma “entidade” “transmite” uma “doença”.

O segundo aspecto a considerar é a forma de representação onde o domínio da ontologia será representado. E neste sentido as ontologias estão apoiadas no conceito de taxonomia, como vimos anteriormente através de Guarino (1998 a).

A taxonomia ou taxionomia surgiu como Ciência das leis da classificação de formas vivas e, por extensão, ciência das leis da classificação. No ambiente dos sistemas

de classificação, das ontologias, da inteligência artificial, é entendida como classificação de elementos de variada natureza. Neste contexto, a utilização de taxonomias permite que se estabeleçam padrões de alto nível para a ordenação e classificação de informação através do uso de mecanismos de herança. O conceito de herança é um dos conceitos mais poderosos no desenvolvimento de software. As máquinas podem compreender corretamente relacionamentos de generalização e especialização entre as entidades atribuindo propriedades às classes gerais e então assumindo que as subclasses herdam estas propriedades. (CAMPOS, M. L. M.; CAMPOS, M. L. A.; CAMPOS, L. M., 2005).

O terceiro componente em nada se assemelha as linguagens documentárias, pois estas não têm por função possibilitarem inferências por sobre o domínio que representam.

No âmbito da Ciência da Informação estudos vêm sendo realizados visando estabelecer bases teóricas e metodológicas para a modelização de domínios de conhecimento. (CAMPOS, 2004). O modo como o núcleo temático de uma ontologia é constituído é de importância capital para a sua elaboração. Esta organização influenciará diretamente nas relações que se estabelecem entre os conceitos, e conseqüentemente nas definições desses conceitos e em sua estruturação no domínio. A partir dos estudos desenvolvidos, verificam-se quatro princípios fundamentais que devem auxiliar o modelizador na elaboração de estruturas conceituais em domínios de conhecimento, como é o caso das ontologias, quais sejam: o método de raciocínio; o objeto de representação; as relações entre os objetos e as formas de representação.

- O método de raciocínio

Modelos vêm sendo elaborados, tradicionalmente, tendo como princípio um dos dois métodos de raciocínio: o método dedutivo, também denominado "top-down", na Ciência da Computação, ou o método indutivo, também denominado "bottom-up", na Ciência da Computação. No âmbito da Ciência da Informação, por sua vez, a Teoria do Conceito (DAHLBERG, 1978b) introduz uma metodologia que poderíamos denominar de híbrida - não só o método dedutivo e não só o método indutivo - mas agregando os dois em um exercício de pensar o particular como um todo e o todo possuindo particulares. Dahlberg, no estabelecimento de sua teoria, apresenta categorias para

representar contextos e, logo depois, analisa os conceitos de um contexto na perspectiva de ordená-los no interior dessas categorias. A princípio, sabe-se da existência das categorias (universais); entretanto, chega-se a elas, a partir de um processo indutivo de análise do conceito.

Na verdade, o que queremos evidenciar é que, antes da utilização do modelo, é necessário ter bastante claro o método utilizado para a sua construção. Numa proposta sistêmica, que leva em conta tanto os princípios de modelização de Le Moigne (1977), quanto à questão da complexidade de Edgar Morin (1990; 2000), acreditamos que tais métodos não são possíveis de serem pensados de forma separada, mas sim de forma sistêmica, onde o indutivo e o dedutivo, a análise e a síntese, são vistos como processos não disjuntos, que se complementam em um todo, ou seja, "análise-síntese-análise-síntese...".

- Objeto de Representação

O segundo princípio diz respeito ao objeto de representação que pode ser definido como a menor unidade de manipulação/representação de um dado contexto.

No âmbito da Ciência da Informação, a Teoria da Classificação (RANGANATHAN, 1967) e a Teoria do Conceito (DAHLBERG, 1978a) se apresentam como teorias que suportam um dos primeiros esforços teórico-metodológicos que evidenciam a questão da unidade de representação. Ranganathan em seus Prolegomena (1967) introduz o conceito de isolado como a unidade mínima e manipulativa de um sistema de classificação. Dahlberg apresenta o Conceito, como unidade mínima, e o define como uma tríade - referente, características e nome. O objeto é o referente, que pode ser classificado como objeto individual, ou geral, que, circunscrito a um dado contexto, requer apropriação de características, sendo-lhe designado um signo lingüístico - um nome.

- As relações entre os objetos

As relações entre os objetos de um dado contexto formam a estrutura conceitual deste contexto e são de natureza diversa. Nesta perspectiva, nos diz Edgar Morin (1990), "todo modelo reflete a complexidade do real". Estas relações foram classificadas segundo Dahlberg (1978 b) em cinco tipos, são elas : relação categorial, relação hierárquica, relação partitiva, relação funcional sintagmática, relação de

equivalência*. A análise será apresentada a partir do que se denomina de movimentos do ato de modelar. Estes movimentos refletem grupos de relações entre conceitos.

O primeiro movimento é a verificação, nas teorias e métodos apresentados, da existência de relações categoriais. Este tipo de relação reúne, em um primeiro grande agrupamento, os objetos por sua natureza, ou seja, entidades, processos, entre outros. Esta relação possibilita, muitas vezes, diminuir erros lógicos no estabelecimento das ligações entre os conceitos, pois determina a natureza do objeto. A noção de Categoria se coloca como um elemento agregador, que reúne os conceitos em um nível mais alto de abstração. Verificado se os objetos estão reunidos ou não por sua natureza, um segundo movimento, não em ordem de precedência, mas de necessidade, é o de verificar como os objetos de mesma natureza se relacionam. Nesta forma de relacionamento determinam-se as relações hierárquicas. Para Dalhberg a relação hierárquica baseia-se em uma relação lógica de implicação, ou seja, nela os conceitos devem ser de mesma natureza, o que não ocorre com o todo e suas partes que, em muitos casos, são de natureza diferentes. Um terceiro movimento do ato de modelar é a análise de como "o objeto se constitui", ou seja, quais são suas partes e elementos. Nesta forma de relacionamento, determinam-se as relações partitivas. Um outro movimento é verificar como objetos de natureza diferente se relacionam e representar esse relacionamento de forma mais consistente, ou seja, a partir da determinação de alguns critérios prescritivos que possibilitam ligações mais criteriosas. Este tipo de relação se apresenta, na Teoria do Conceito, como Relação Funcional Sintagmática, ou seja, relações que se estabelecem entre categorias. As relações funcional-sintagmáticas, diferentemente das relações de natureza paradigmáticas (hierárquicas e partitivas), podem ser reconhecidas como relações que tornam evidente uma dada demanda, ou função, entre os objetos no mundo fenomenal. Por último, é importante verificar um dado tipo de relação que não mais se constitui entre conceitos, mas entre a forma de expressar os conceitos, ou seja, se dá no âmbito da língua: a chamada Relação de Equivalência que evidenciam relacionamentos de sinonímia ou quase-sinônimos.

Podemos afirmar que um corpo de conhecimento representado é baseado em uma conceituação, dos objetos, dos conceitos e de outras entidades, que se supõem

* Um estudo mais aprofundado dos tipos de relações, apresentando também comparações entre teorias da Ciência da Informação e da Computação pode ser encontrado em Campos (2004).

existir em alguma área de interesse e os relacionamentos que os ligam. Desta forma, toda base de conhecimento é comprometida com alguma conceituação. Esta deve poder ser representada para atender as necessidades de sistemas automatizados.

Nestes sistemas, são os mecanismos de representação de conhecimento que podem facilitar quais processos de formalização sobre os objetos e suas relações em contextos pré-definidos podem ser facilmente representados. No âmbito das ontologias estes mecanismos permitem a sistematização dos conceitos e conseqüentemente a elaboração de definições consistentes, visando possibilitar futuras inferências sobre o domínio. As bases teóricas e metodológicas, fruto de anos de pesquisa no âmbito do tratamento e recuperação da informação, podem auxiliar esta nova ferramenta de organização de conhecimento que são as ontologias.

- As formas de Representação Gráfica

Um modelo conceitual deve ser visto, também, como um espaço comunicacional, onde transpomos o mundo fenomenal para um espaço de representação. Desta forma, consideramos importante investigar formas gráficas de representação, porque, em alguns casos, estas formas "inibem" as possibilidades representacionais. A área da Ciência da Informação, apesar de ter teorias bem fundamentadas sobre o conceito e relações conceituais, é fraca em modelos que auxiliem a elaboração de representações gráficas. Tanto na Teoria da Classificação Facetada como na Teoria do Conceito, não é apresentado nenhum modelo para expressar graficamente as relações conceituais. Por outro lado, apesar da Ciência da Computação avançar nesta perspectiva, ela ainda está refém de possibilidades representacionais que estão aquém das ontologias, pois estas últimas necessitam de representações gráficas que possam agregar e diferenciar os diversos tipos de relações entre os conceitos. Destacamos que as representações gráficas carecem muito mais que qualquer outro ponto analisado, de um espaço maior de investigação.

3.2. A modelização de domínios de conhecimento na Ciência da Computação

Pouco a pouco na história do desenvolvimento da computação a ênfase no processamento de dados se inverte. Na década de 60 surgem às primeiras bases e bancos

de dados, e o conceito de sistemas de informação – pouco processamento sobre uma quantidade muito grande de dados. Essa ênfase em dados, que podem se transformar em informação, o coloca a questão da sua semântica. Sistemas computacionais também vêm se tornando mais e mais complexos ao longo da sua evolução; surge também a necessidade de planejar sistematicamente sistemas computacionais.

A modelagem de bases de dados (CHEN, 1976) e de sistemas computacionais em geral (UML) surge para garantir que eles incluam as informações e funcionem conforme previsto por seus usuários. Uma característica dessas abordagens é separar claramente a construção de um modelo abstrato – a modelagem conceitual e semântica – da posterior implementação computacional. Metodologias de modelagem de sistemas como a UML – Unified Modeling Language* – incluem a especificação de interfaces homem-máquina e os chamados casos de uso, mostrando a necessidade, para seu bom funcionamento, de visualizar sistemas computacionais dentro das complexas relações sociais onde estão inseridos.

Como se vê, na modelagem de sistemas, as abordagens meramente computacionais, algorítmicas, baseadas nos fluxogramas de programas, logo mostraram suas limitações. Um sistema computacional é muito mais que um algoritmo. Pierre Lévy (1993, p. 19) chama a atenção para as interações humano-tecnológicas nas redes de informação atuais como os verdadeiros *sujeitos do pensamento*. “*Os coletivos cosmopolitas compostos de indivíduos, instituições e técnicas não são somente meios ou ambientes para o pensamento, mas sim seus verdadeiros sujeitos.*”. Sistemas computacionais, com suas interações humano-máquina e humano-humano, mediadas por computadores e redes, por sua vez inseridas em complexas interações sócio-culturais, são sistemas altamente complexos, recursivos, em que o modelador é também parte do sistema, influencia e é influenciado por ele (LE MOIGNE, [s.d.]).

Neste sentido a modelagem de sistemas computacionais, entre eles as ontologias, vai na direção do preconizado por Hjørland (2002), e já investigada por pesquisadores da Ciência da Computação como Prieto-Diaz (1990) e Berzittis (1999), de uma visão *sócio-cognitiva*, de *análise de domínios*. Dado que computadores trabalham somente seguindo instruções, dentro de um formalismo rigoroso, para que se possa usufruir dos benefícios de representar, simular e avaliar o comportamento de domínios no espaço

* <http://www.uml.org/>

computacional, existe a necessidade de formalizar o mais completamente possível esses domínios.

4. Conclusões

Existem vários caminhos para endereçar as questões colocadas aqui para o desenvolvimento de ontologias. Por um lado, se poderia caminhar-se para uma padronização das linguagens para representá-las, como o OWL. Outro caminho seria referenciar ontologias umas às outras, complementando-as, e cada ontologia específica à ontologias chamadas de alto nível, contendo conceitos mais genéricos e consensuais, com é explicitamente a proposta da Suggested Upper Merged Ontology (SUMO), <http://www.ontologyportal.org/>, ou de propostas como Basic Formal Ontology (BFO), <http://www.ifomis.org/bfo>, Cyc, <http://www.cycfoundation.org/>, WordNet - a lexical database for the English language -, <http://wordnet.princeton.edu/>. Poder-se-ia ainda fazer todo um trabalho de informação no sentido de sistematizar e montar registros e bibliotecas de ontologias, prevendo sua padronização, a não duplicação de esforços, seu re-uso.

Uma outra linha de intervenção seria trabalhar em metodologias para o desenvolvimento de ontologias, uma área de pesquisa muito significativa, também conhecida como “ontology engineering” (DEYEDZIC, 2002). Sendo o esqueleto básico de uma ontologia é uma taxonomia, Smith (2001) traça uma linha evolutiva desde simples vocabulários controlados até ontologias formais processáveis por programas, que incluem regras lógicas para restringir o significado de relações e propriedades. Nesse aspecto a Ciência da Informação tem uma contribuição significativa a dar.

A pesquisa em ontologia hoje recoloca velhos problemas para as práticas de gestão de conhecimentos que vêm sendo utilizadas pela humanidade até hoje. Como adquirir conhecimento, como torná-lo consensual, validá-lo de uma forma consistente, para que ele possa ser re-utilizado? Como representa-lo e registrá-lo de forma a que ele possa ser re-utilizado? Como compatibilizar acervos de conhecimento; como somar suas sinergias? Quais os mecanismos sociais para isso? Novas formas tem sido tentadas, como indica D’Ascenzo (2004) para o processo de sequenciamento genômico. Essas questões, devido a sua abrangência, só podem ser enfrentadas a partir de uma ampla

colaboração de saberes, dentro de uma visão interdisciplinar, para a qual a Ciência da Informação sempre esteve atenta.

Para que a Ciência da Informação possa contribuir com todo o seu potencial e experiência para o desenvolvimento da Web Semântica é necessário que a CI faça um movimento de se aproximar destas questões, de modo a compreendê-las mais claramente e qual poderia ser sua contribuição. Porém esta possível cooperação não é dada nem esta pronta. É necessária uma aproximação, engajamento em projetos comuns, claramente de caráter interdisciplinar, testar as metodologias no uso comum para que possam ser aperfeiçoadas e se tornem operacionais.

A Web e, em especial, o projeto Web Semântica, mais do que avanços tecnológicos, se constituem em marcos culturais profundos. Uma das questões mais sérias aí colocadas é a questão do conhecimento, dos novos modos de conhecer, dos novos modos de armazenar, preservar e ascender ao conhecimento. É neste cenário, balizado pelo surgimento do computador, as chamadas ciências da concepção – a tanto a Ciência da Computação quanto a Ciência da Informação, conforme Gómez (2003). A Ciência da Informação vem de uma origem claramente marcada pelo uso do computador para o processamento de informações, vem evoluindo na direção de enxergar os fluxos de informação dentro de contextos sociais mais amplos. Essa parece ser também a trajetória da Ciência da Computação. Como coloca Guarino (1995): “ontologia e modelagem conceitual necessitam ser estudadas a partir de uma perspectiva altamente interdisciplinar”*. Essa sempre foi a perspectiva da Ciência da Informação. Assim colocada, essa perspectiva faz antever um feliz encontro das duas disciplinas.

REFERÊNCIAS

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. *The semantic web*. Scientific American, May, 2001. Disponível em <<http://www.scian.com/2001/0501issue/0501berners-lee.html>>, Acesso em 24 maio 2001.

BERZITISS, A. T. Domain analysis for business software systems. *Information Systems*, v. 24, n. 7, p. 555-568, 1999.

* Tradução nossa

- BORGES, Jorge Luis. *Ficciones*. Madrid: Alianza Editorial, 2007.
- BRETON, Philippe. *História da informática*. São Paulo: Ed. UNESP, 1991. 260 p.
- BROOKES, B. The foundations of information science - Part I. Philosophical aspects. *Journal of Information Science*, v. 2, p. 125-133, 1980.
- BUCHANAN, Bruce G.; SMITH, Reid G. Fundamentals of expert systems. *Annual Review of Computer Science*, n. 3, p. 23-58, 1988. Disponível em: <arjournals.annualreviews.org>. Acesso em: 20 abr. 2008.
- CAMPANARIO, J. M. El sistema de revisión por expertos (peer review): muchos problemas y pocas soluciones. *Revista Española de Documentación Científica*, v. 25, n. 3, 2002. Disponível em: <http://redc.revistas.csic.es/index.php/redc/article/view/107/171>. Acesso em: 21 abr. 2008.
- CAMPANARIO, J. M. Peer Review for journals as it stands today - Part 1. *Science Communication*, v. 19, n. 3, p. 181-211, mar. 1998. Disponível em: <http://www2.uah.es/jmc/ai4.pdf >. Acesso: em 21 abr. 2008.
- CAMPOS, M. L. A. INTEGRAÇÃO DE ONTOLOGIAS: O domínio da Bioinformática e a problemática da compatibilização terminológica. (Projeto de Pesquisa submetido ao CNPq no período de 2005 a 2008). Universidade Federal Fluminense- Departamento de Ciência da Informação, 2005a.
- CAMPOS, M. L. A *Linguagem documentária*: teorias que fundamentam sua elaboração. Niterói: EDUFF, 2001.
- CAMPOS, M. L. M.; CAMPOS, M. L. A.; CAMPOS, L. M. . Web semântica e a gestão de conteúdos informacionais. In: Carlos H. Marcondes; Hélio Kuramoto; Lídia Brandão Toutain; Luís Sayão. (Org.). *Bibliotecas digitais: saberes e práticas*. Salvador, EDUFBA; Brasília: IBICT, 2005, p. 55-75.
- CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação dos princípios fundamentais. *Ciência da Informação*, Brasília, v. 33, n. 1, jan./abr., p. 22-32, 2004.
- CAMPOS, M. L. A. Integração de Ontologias: o domínio da bioinformática. *RECIIS - Revista de Comunicação, Informação; Inovação em Saúde*, v. 1, p. 117-121, 2007.
- CASTELLS, M.. *A sociedade em rede*. Vol. 1 - A era da informação: economia, sociedade e cultura. São Paulo: Paz e Terra, 1999.
- CHEN, Peter Pin-Chan. The entity relationship model: towards a unified view of data. *ACM transactions on database systems*. v. 1, n. 1, p. 9-36, mar. 1976.
- CORAZZON, R. Ontology: a resource guide for philosophers. 2000. Disponível em: <<http://www.formalontology.it>>. Acesso em: 01 jul. 2006.
- DAHLBERG, I. A Referent-oriented analytical concept theory of interconcept. *International Classification*, Frankfurt, v.5, n.3, p.142-150, 1978a.
- DAHLBERG, I. *Ontical structures and universal classification*. Bangalore: Sarada Ranganthan Endowment, 1978b. 64 p.
- D'ASCENZO, M. D.; COLMER, A.; MARTIN, G. B. PeerGAD: a peer-review-based and community-centric web application for viewing and annotating prokaryotic genome

sequences. *Nucleic Acids Research*, v. 32, n. 1, p. 3124-3135, 2004. Disponível em: <<http://nar.oxfordjournals.org/cgi/content/abstract/32/10/3124> > . Acesso em: 21 abr. 2008.

DE BRUIJN, Jos; EHRIG, Marc; FEIER, Cristina; MARTÍN-RECUERDA, Francisco; SCHARFFE, François; WEITEN, Moritz. Ontology Mediation, Merging, and Aligning. In: DAVIES, John Davies; STUDER, Rudi; WAREN, Paul (ed.). *Semantic web technology: Trends and research in ontology-based systems*. Chichester (England): John Wiley & Sons, 2006. p. 95-112.

DERVIN, B. An overview of sense-making research: concepts, methods and results. Paper presented at the annual meeting of the *International Communication Association*. Dallas (Texas), 1983.

DEYEDZIC, V. Understanding ontological engineering. *Communications of the ACM*, v. 45, n. 4, apr. 2002.

DING, Yin; FOO, Schubert. Ontology research and development. Part 1 - a review of ontology mapping and evolving. *Journal of Information Science*, v.28, n.10, p. 123 - 136, 2002.

DING, Yin; FOO, Schubert. Ontology research and development. Part 2 - a review of ontology generation. *Journal of Information Science*, v.28, n.4, p. 375-388, 2002.

DUPUY, Jean-Pierre. *Nas origens das ciências cognitivas*. São Paulo: UNESP, 1996. (Biblioteca Básica).

FALBO, R. A. Integração de conhecimento em um ambiente de desenvolvimento de software. 1998. Tese de doutorado (COPPE/Universidade Federal do Rio de Janeiro, 1998. 81p.

FARRADANE, J. Relational indexing - Part I. *Journal of Information Science*, v. 1, p. 267-276, 1980.

FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. **Proceedings** of the AAAI1997 Spring Symposium on Ontological Engineering, 1997. p. 33-40.

FERNÁNDEZ-LÓPEZ, M. Overview of methodologies for building ontologies. In: IJCAI-99 Workshop on ontologies and problem-solving methods: lessons learned and future trends. *Intelligent Systems*, 2001, p. 26-34. Disponível em: <www.Isi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>.

FORSYTHE, D. E.; BUCHANAN, B. G. Knowledge acquisition for expert systems: some pitfalls and suggestions. *IEEE Transactions on Systems, Man and Cybernetics*, v.19, n. 3, p-435-442, may/jun. 1989. Disponível em: <<http://ieeexplore.ieee.org/iel1/21/1336/00031050.pdf?tp=&isnumber=&arnumber=31050>>. Acesso em: 21 abr. 2008.

GANGEMI, A.; PISANELLI, D. M.; STEVE, G. An Overview of the ONIONS Project: applying ontologies to the integration of medical terminologies. **Data Knowledge Engineering**, v. 31, n. 2, p.183-220, 1999.

- GARDIN, Jean-Claude. Vers un remodelage des publications savantes: ses rapports avec sciences de l'information. In: COLLOQUES ISKO-FRANCE. Filtrage et résumé automatique de l'information sur les réseaux., Conference invitee, Université de Nanterre - Paris X, 2001, Proceedings... (Conference proceedings).
- GENARO, Sergio. *Sistemas especialistas: o conhecimento artificial*. Rio de Janeiro: Livros Técnicos e Científicos Editora, 1985.
- GOMEZ, Maria Nélide González de. Comunicação-informação-cognição: interfaces. In: INFORMAÇÃO, COMUNICAÇÃO, COGNIÇÃO: na complexidade dos sistemas de saúde. Rio de Janeiro: FIOCRUZ, 2003.
- GOMEZ, M. N. G. O papel do conhecimento e da informação nas formações políticas ocidentais. *Ciência da Informação*, Brasília, v.16, n.2. p.157-167. jul./dez. 1987.
- GRUBER, T.R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, p. 199-220, 1993.
- GRÜNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing at the 14th International Joint Conference on Artificial Intelligence IJCAI1995*, 1995.
- GUARINO, N. Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human Computer Studies*, special issue on formal Ontology, conceptual analysis and knowledge representation. 1995. Disponível em: <<http://www.ldv.uni-trier.de:8080/ldvpage/naumann/Ontologie/GUARINO/FORMONTKR.PDF>>. Acesso em: 30 dez. 2005.
- GUARINO, N. Semantic matching: formal ontological distinction for information organization, extraction, and integration. In: PAZIENZA, M. T. (Ed.) *Information Extraction: a multidisciplinary approach to na emergig information technology*, 1997.
- GUARINO, N. Formal ontology and information systems. In: FOIS '98, 1., 1998, Trento, Italy. *Proceedings...* Amsterdam: IOS Press; Tokyo: Omsa, 1998 a. p. 3-15.
- _____. Some ontological principles for designing upper level lexical resorces. In: International Conference on Language Resources and Evaluation, 1., 1998, Granada. *Proceedings...*1998 b. Disponível em: <http://www.loa-cnr.it/Papers/LREC98.pdf>. Acesso em: 15 out. 2005.
- GUARINO, N.; WELTY, Christopher. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, v. 45, n. 2, p. 61-65, 2002.
- _____. Towards a methodology for ontology based model engineering. In: ECOOP-2000 Workshop on Model Engineering, 2000, *Proceedings...* Disponível em <ECOOP-2000 Workshop on Model Engineering, 2000>. Acesso em 03 jun. 2008.
- HJØRLAND, Birger. Domain analysis in information science: Eleven approaches - traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p. 422- 462, 2002.

- HOROBIN, D. F. The philosophical basis of peer review and the suppression of innovation. **Journal of the American Medical Association**, v. 263, n.10, mar. 9, 1990. Disponível em: <<http://jama.ama-assn.org/cgi/content/abstract/263/10/1438>>. Acesso em: 21 abr. 2008.v. 263, n. 10, 1990.
- JACOB, Elin K. Ontologies and the Semantic Web. *Bulletin of the American Society for Information Science and Technology*, April/May, 2003. p. 19-22.
- KAJIKAWA, Y.; ABE, K.; NODA, S. Filling the gap between researchers studying different materials and different methods: a proposal for structured keywords. *Journal of Information Science*, v. 32, p. 511-524, 2006.
- KHOO, C.; NA, J. C. Semantic Relations in Information Science. *Annual Review of Information Science and Technology*, p. 157-228, 2007.
- LE MOIGNE, J-L. *A teoria do sistema geral: teoria da modelização*. Lisboa: Instituto Piaget, 1977.
- LE MOIGNE, J.-L. *Formalisms of systemic modeling*. 1992. Disponível em: <<http://www.mcxapc.org/docs/ateliers/0505formalismes.pdf>> . Acesso em: 05 mar. 2005.
- LUHN, H., Keyword in Context Index for Technical Literature. *American Documentation*, v. 11, n. 4, p. 288-295, 1960.
- MAGALHÃES, Gildo. Das máquinas de calcular à informática. *Rev. SBHC*, n. 17, p. 21-28, 1997.
- MARCONDES, Carlos. Henrique. Representação e economia da informação. *Ciência da Informação*, Brasília, v. n. p. 2000. Disponível em <http://www.scielo.br/scielo.php?pid=S0100-19652001000100008&script=sci_arttext>. Acesso em 11 abr. 2008.
- MARCONDES, C. H; MENDONÇA, M. A. R; MALHEIROS, L. R. Uma proposta de modelo de representação do conhecimento contido no texto de artigos científicos publicados na Web em formato legível por programas. *DatagramaZero*, 2006. Disponível em <http://www.dgz.org.br/out06/F_I_aut.htm>.
- METAXIOTIS, K.; PSARRAS, J. E.; ASKNOUNIS, D. Building ontologies for production scheduling systems: towards a unified methodology. *Information management & computer security*, v. 9, n. 1, p. 44-50, 2001.
- MORIN, Edgar. *Introdução ao pensamento complexo*. 2.ed. Lisboa: Instituto Piaget, 1990.
- MORIN, E.; LE MOIGNE, J-L. *A inteligência da complexidade*. São Paulo: Petrópolis, 2000.
- OTLET, Paul. *Traité de documentation: le livre sur le livre. Théorie et pratique*. Liège: Centre de Lecture Publique de la Communauté Française de Belgique, 1989. 432p.
- PASLARU-BONTAS, Elena. A Contextual approach to ontology reuse: methodology, methods and tools for the semantic web. 2007. Dissertation (Fachbereich Mathematik und Informatik, Freien Universität Berlin).
- PERREAULT, J. Categories and relators. *International Classification*, Frankfurt, v. 21, n. 4, p. 189-198, 1994.

- PESSIS-PASTERNAK, Guitta. *Do caos à inteligência artificial: quando os cientistas se interrogam*. São Paulo: UNESP, 1993.
- PRIETO-DIAZ, R. Domain analysis: an introduction. *SIGSOFT Software Engineering Notes*, v. 15, n. 2, p47-54, 1990.
- RANGANATHAN, S. R. *Prolegomena to library classification*. New York: Asia Publishing House, 1967.
- SALES, L. F.; CAMPOS, M. L. A.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais e sua aplicação. In: **Anais...** Marília, 7ENANCIB, 2006, Marília, 2006.
- SARACEVIC, T. A natureza interdisciplinar da ciência da informação. *Ciência da Informação*, Brasília, v. 24, n. 1, 1995.
- SHERA, J. H., Kent, A; PERRY, J. W. (Ed.). *Information systems in documentation*. New York: Interscience, 1957. *Advances in Documentation and Library Science*, v. 1.
- SMITH, B. *Ontology and information systems*. 2002. Disponível em: <[http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf)>. Acesso em: 26 maio 2008.
- SMITH, B.; CEUSTERS, W.; KLAGGES, B.; KÖHLER, J.; KUMAR, A.; LOMAX, J.; MUGALL, C.; NEUHAUS, F.; RECTOR, A.; ROSSE, C. Relations in biomedical ontologies. *Genome Biology*, v. 6, n. 5. 2005. Disponível em: <<http://genomebiology.com/2005/6/5/R46>>. Acessado em: 4 dez. 2005.
- SMITH, B.; KOHLER, J.; KUMAR, A. On the Application of formal principles to life science data: a case study in the gene ontology. *Lecture Notes in Computer Science*, v. 2994, p. 79-94, 2004. Disponível em: <http://ontology.buffalo.edu/medo/Database_Integration.pdf>. Acesso em: 01 ago. 2007.
- SMITH, B.; KÖHLER, J.; MUNN, K.; RÜEGG, A.; SKUSA, A. Quality control for terms and definitions in ontologies and taxonomies. *BMC Bioinformatics*, v. 7, n. 212, 2006. Disponível em: <<http://www.biomedcentral.com/1471-2105/7/212>>. Acesso em: 8 abr. 2008.
- SMITH, B.; WELTY, C. Ontology: Towards a New Synthesis. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SCIENCE, Ogunquit, Maine, USA, 2001, **Proceedings...** New York: ACM, 2001. p. 3-9. Disponível em: <<http://www.cs.vassar.edu/~weltyc/papers/fois-intro.pdf>>. Acesso em: 09 maio 2008.
- SORGEL, Dagobert. The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*, v.5, n.12, p.1119 – 1120, 2000.
- SURE, Y.; STAAB, A.; STUDER, R. Methodology for development and employment of ontology based knowledge management applications. *SIGMOD Record*, v. 31, n. 4, p. 18-23, dec. 2002. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2002_sigmod-methodology.pdf>.
- SWARTOUT, W. ; TATE, A. Guest editors' introduction: ontologies. *IEEE Intelligent Systems*, v. 14, n.1, p. 18-19, Jan. 1999.
- TENÓRIO, Robinson. *Computadores de papel: máquinas abstratas para um ensino concreto*. 2. ed. São Paulo: Cortez, 2001. (Coleção Questões de nossa época, 80).

THE SEMANTIC WEB. *Bulletin of The American Society for Information Science and Technology*, v. 29, n. 4, apr./may 2003. (Special Section).

USCHOLD, M.; KING, M. Towards a methodology for building ontologies. *Proceedings of the IJCAI1995*, Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.

VELTMAN, K. H. Towards a semantic web for culture. *Journal of Digital Information*, v 4, n. 4, 2004. Disponível em: <<http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/>>. Acesso em: 29 jun. 2007.

VICKERY. B. C. Ontologies. *Journal of Information Science*, v. 23, n. 277, p. 277-286, 1997.

WEINSTEIN, Peter C. Ontology-Based Metadata: transforming the MARC Legacy. *Digital Libraries*, Pittsburg, p. 254-263, 1998.

WILSON, T.D. The Work of the British Classification Research Group. In: WELLISCH, H. (ed.) *Subject retrieval in the seventies*. Westport: Greeword, 1972. p. 62-71.