

Do texto às ontologias: uma perspectiva para a ciência da informação

Marcelo Schiessl

Doutorando em ciência da informação pela Faculdade de Ciência da Informação da Universidade de Brasília. Brasília, DF – Brasil.

E-mail: schiessl@ig.com.br

Marisa Bräscher

Doutora em ciência da informação pela Universidade de Brasília, UnB, Brasília, DF. Professora da Universidade Federal de Santa Catarina.

Departamento de Ciência da Informação. Florianópolis, SC – Brasil.

E-mail: marisa.brascher@gmail.com

A formal definition of ontology is presented by utilizing Logic and Mathematics, which are essential for automation of computerized procedures. Some models proposed are interrelated for construction of ontology and consolidation of a vision which can be used in Information Systems. Arguments are presented for the automatic construction of ontology from the text as a way of making the production of ontology feasible and as fast as required for the present days and so providing a quality to information structures spread out in the web digital world.

Keywords

Ontology. Semantic Web. Ontology learning. Text mining. Information system.

Resumo

Discute o termo “ontologia” sob a perspectiva da filosofia e da ciência da informação. Apresenta as raízes que constituíram o termo e a relação com o uso atual entre pesquisadores de áreas diversas, mas relacionado ao contexto computacional. Mostra uma definição formal de ontologia que utiliza recursos da lógica e da matemática essenciais para automatização de procedimentos computadorizados. Relaciona alguns modelos propostos para construção de ontologia e os consolida em uma visão que pode ser utilizada em sistemas de informação. Argumenta sobre a construção automática de ontologias a partir de textos como forma de viabilizar a produção de ontologias na velocidade requerida dos dias atuais e, assim, fornecer qualidade às estruturas de informações dispersas no mundo digital da Web.

Palavras-chave

Ontologia. Web semântica. Aprendizado de ontologia. Mineração de textos. Sistema de informação.

From the text to ontology: a perspective for information science

Abstract

The objective of this paper is to discuss the term “ontology” with reference to Philosophy and Information Science. The rationale which constitutes the term is analyzed in relation to the present use by researchers of different areas, but always related to the computer context.

INTRODUÇÃO

A informação é tão valiosa quanto qualquer outro bem. A velha máxima – informação é poder – representa os dias atuais de maneira bastante apropriada. Ela se tornou um bem de consumo tão valorizado quanto necessário, de acordo com o ineditismo ou a relevância. Nas sociedades mais desenvolvidas, o comércio da informação aponta a vanguarda do comportamento humano e uma nova necessidade, a instantaneidade informacional. Entretanto, vive-se na era da informação digital, cujo volume supera a capacidade de gestão e absorção humanas. Tal situação obriga a humanidade a buscar auxílio na inventividade e na capacidade de superação. Essa ajuda, hoje, é conseguida com a interação homem-máquina.

A tecnologia é uma parceira a serviço da informação. Dispõe-se de inúmeros recursos automatizados construídos com o propósito de tornar a informação acessível. A ciência da informação (CI) se ocupa em entender a natureza e o uso da informação. Essas atividades possuem interseções

com outras áreas da ciência, em especial a ciência da computação (CC). O desenvolvimento de artefatos tecnológicos auxilia na tarefa de gerenciamento de informação. Eles apoiam quase todo processo de estruturas tecnológicas – como repositórios para armazenamento, redes de comunicação – aos *softwares* especializados que visam representar o conhecimento humano.

Mais uma fronteira é vencida pela globalização digital. A rede mundial de computadores, ou simplesmente Web, possibilitou a aceleração da disseminação da informação, dado que ela extrapola as fronteiras entre países. Assim, o processamento e a interpretação compartilhados das informações em repositórios na Web tem sido o foco de muitas pesquisas que visam à adaptação das máquinas ao comportamento humano.

A representação do conhecimento humano é peça chave. Nessa perspectiva, a interpretação semântica da informação não pode ser negligenciada. Para tanto, buscam-se maneiras de representar o conhecimento humano de modo que seja processável por computadores, como fazem as ontologias.

Contudo, a construção de ontologias é, notadamente, complexa e consumidora de tempo. Isso por que elas são tipicamente compartilhadas por um grupo ou comunidade de pessoas que, frequentemente, divergem quanto às definições e conceitualizações em determinados domínios. Além disso, o custo pode se mostrar elevado, pois a manutenção de especialistas de domínio é dispendiosa.

Automatização total ou parcial é uma alternativa viável, se aplicada ao processo de construção de ontologias. Isso pode significar redução de custos, visto que o acúmulo de informações digitais representa uma fonte rica de conhecimento a ser apreendido e representado na forma de ontologia. Não se trata de substituição do homem, mas da realocação dele para tarefas intelectuais que, até então, não estão acessíveis às máquinas.

ONTOLOGIA

A ciência procura alternativas nos sistemas de representação do conhecimento. Na busca de capacitação das máquinas ao reconhecimento e interpretação de informações, o homem se depara com uma questão filosófica secular: o que é o conhecimento e como adquiri-lo? A área de inteligência artificial (IA) tem devotado esforços para desenvolver sistemas capazes de “copiar” o raciocínio humano, isto é, de adquirir novo conhecimento a partir de bases de conhecimento (CIMIANO, 2006).

No princípio era o termo. Historicamente, o termo ontologia tem origem na filosofia que Aristóteles chamou de Filosofia Primeira. Ela estuda o “Ser” enquanto ser. Em grego, o termo onto significa “Ser”, enquanto logia, estudo ou conhecimento. Assim, ensina Chauí (2003) que ontologia significa o estudo dos entes, das coisas como verdadeiramente são, isto é, do “Ser”.

O pensamento filosófico sustenta o conceito. É interessante observar que o filósofo Platão, mestre de Aristóteles, lançou as bases para o que atualmente se entende por ontologia. Ele afirmou que existem dois mundos: o mundo das ideias, que seria o mundo sem imperfeições, enquanto o mundo do sensível seria este em que vivemos e os objetos observados seriam apenas sombras ou representações imperfeitas das ideias ou formas. Dessa abordagem, é razoável supor que os objetos são os entes verdadeiros e perfeitos, e as sombras, as projeções desses objetos com as imperfeições que os modelos carregam em si.

Mas, afinal, o que é ontologia para a CI? Numa proposta, Gruber (1993) afirma que ontologia é a especificação formal e explícita de conceitualização. Borst (1997) complementa que essa conceitualização deve ser compartilhada, na qual os grupos de objetos e suas relações são refletidas por vocabulário representacional. Noutra, Guarino (1995) acrescenta que o estudo dela pode beneficiar os processos de

construção de conhecimento produzindo bases de conhecimento de alta qualidade. Noy e Hafner (1997) entendem que um corpo de conhecimento formalmente representado está baseado em conceitualização. Essa representa um grupo de objetos, conceitos ou outras entidades sobre os quais o conhecimento e suas relações são expressos. Já para Hovy (2002), ela é um conjunto de termos associados com definições em linguagem natural que utiliza relações formais e é relativo a algum domínio de interesse. Em visão mais recente, que não esgota o assunto, Daconta et al. (2003) asseguram que ontologia define termos e conceitos comuns na representação e descrição de determinada área de conhecimento e, assim, padroniza o significado.

Infere-se daí que ontologia como explicação sistemática da própria existência se estende para o domínio de sistemas de informação (SI). De tal maneira, pode-se descrever uma ontologia¹ pela definição de um grupo de termos representativos de determinado domínio. A tarefa desse domínio é representar formalmente essa existência.

Nesse contexto, ontologia é um modelo que reflete uma visão de mundo. Depreende-se das argumentações anteriores que o objetivo dela é delimitar a quantidade de interpretações e caracterizar os significados básicos de categorias utilizadas para descrever determinado domínio. Logo, pragmaticamente, tem-se uma representação limitada e incompleta, porém útil, do domínio de interesse.

DEFINIÇÃO

Atualmente, o termo ontologia é comumente utilizado, na CI, com sentido definido por Gruber (1993) e Borst (1997), que pode ser visto como fonte de recursos que representa um modelo

¹ Observe que o termo ontologia no contexto prático da representação do conhecimento é escrito com letra minúscula. Apesar do empréstimo do nome da filosofia, o propósito é mais humilde que a explicação da própria existência.

conceitual de determinado domínio, descrito por vocabulário compartilhado por um grupo ou comunidade.

Do ponto de vista formal, de acordo com Shamsfard e Barforoush (2003) e Cimiano, (2006) uma ontologia pode ser definida como uma estrutura:

$$O = \{C, R, A, Arquilexema\} \quad (1)$$

que consiste em:

C – é o conjunto não vazio de conceitos, incluindo relação de conceitos e o arquilexema;

R – é o conjunto de todas as afirmações em que dois ou mais conceitos são relacionados entre si. R é particionado em dois subgrupos H e N, nos quais H representa o conjunto de todas as afirmações em que a relação é taxonômica, isto é, hierárquica e N é o conjunto de todas as declarações em que a relação não é taxonômica, ou seja, não guarda relação hierárquica entre os conceitos;

A – é o conjunto de axiomas ou atributos de restrições;

Arquilexema – ² é o conceito mais abrangente na abordagem hierárquica.

A formalização matemática expressa na equação simboliza uma ontologia e os elementos que a compõem. Isso é importante para que se possa traduzir linguagem natural em artificial, pois a última pode ser compreendida por máquinas.

² Tradução da palavra *Top* em inglês. Segundo Houaiss (S.D): a palavra ou a locução que reúne o conjunto de traços semânticos (semas) pertinentes e comuns a diversas unidades de uma série de palavras de um mesmo campo semântico, e representa a neutralização da oposição multilateral de traços semânticos específicos das unidades dessa série, p.ex.: recipiente é arquilexema da série vaso, vasilha, taça, tigela, caixa etc. Certos arquilexema são criados em vocabulários especializados, p.ex.: eletrodoméstico, hortigranjeiro ou científico, p.ex.: o suf. nom. -ase que significa “enzima” é arquilexema para a série diástase, isomérase, polimérase.

CONSTRUÇÃO

A construção de ontologias tem sido o tema principal de várias atividades de pesquisa que objetivam utilizá-las em sistemas de informação (SI). O desafio está no gargalo da aquisição do conhecimento e no grande consumo de tempo para construí-las e integrá-las a outros domínios e aplicações.

Recentemente, as ontologias são reconhecidas como componente importante para a construção da Web Semântica. Segundo Berners-Lee et al. (2001), ela não é uma Web apartada da atual, mas uma extensão. Nela, a informação está bem definida, o que possibilita o trabalho cooperativo entre computadores e pessoas. Para tanto, máquinas devem ter acesso a conjuntos de informações estruturadas e a grupos de regras de inferência que podem conduzir ao raciocínio automatizado.

Numa visão com viés da tecnologia da informação³, Maedche e Staab (2001) declaram que ontologias servem como esquemas de metadados que fornecem vocabulários de conceitos com semânticas explicitamente definidas e processáveis por máquinas. Nessa visão, as ontologias poderiam ser construídas a partir da especificação de vocabulário compartilhado entre especialistas de domínio ou com reutilização de ontologias disponíveis.

É consenso entre pesquisadores de que não há a melhor maneira de se construir uma ontologia. A CC atua fortemente nas pesquisas da área e direciona alguns passos nessa tarefa. Assim, as etapas de construção de ontologias têm sido muito influenciadas pelas atividades de engenharia de *software*.

De acordo com Pinto e Martins (2004), as metodologias mais representativas na construção de ontologias são TOVE, ENTERPRISE e METHONTOLOGY⁴, sendo esta fortemente influenciada pela metodologia Rational Unified Process (RUP), bem conhecida dos desenvolvedores de *software*. Uma consolidação de várias metodologias para representar o ciclo de vida do desenvolvimento de ontologias é apresentada na figura 1.

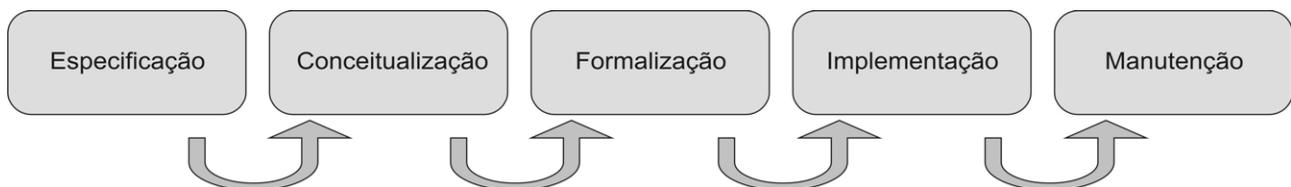
Cada uma dessas atividades é comentada a seguir:

ESPECIFICAÇÃO – Identifica o propósito e o âmbito da ontologia;

CONCEITUALIZAÇÃO – Descreve, em modelo conceitual, a ontologia a ser construída, de forma que atenda às especificações do passo anterior. O modelo conceitual de ontologia consiste no domínio de conceitos e as relações entre eles. As relações reforçam as conexões mais fortes entre grupos de conceitos. Os grupos de conceitos fortemente relacionados geralmente correspondem a diferentes módulos (subontologias) em que o domínio pode ser decomposto;

FIGURA 1

Ciclo de vida de desenvolvimento de uma ontologia



³ Essa afirmação apresenta a percepção de estudos oriundos da CC, que não necessariamente é compartilhada pelos autores deste trabalho. Entretanto, achamos importante apresentar diversos pontos de vista.

⁴ Não serão detalhadas as metodologias citadas, pois fogem ao escopo deste artigo. Ao leitor interessado, sugere-se a leitura de Pinto e Martins (2004), na qual se faz uma comparação entre elas.

FORMALIZAÇÃO – Transforma a descrição conceitual em modelo formal, isto é, a descrição do domínio no passo anterior é representada em linguagem formal, ainda que não seja a forma final. Conceitos são normalmente definidos através de axiomas que delimitam as interpretações possíveis para o significado desses conceitos. Conceitos são geralmente organizados hierarquicamente através de uma relação estruturante, tal como “é-um” (classe-superclasse, instância-classe) ou “parte-de”;

APLICAÇÃO – Implementa a ontologia formalizada em linguagem de representação de conhecimento. Para isso, escolhe-se uma linguagem para representação e escreve-se o modelo formal na linguagem escolhida;

MANUTENÇÃO – Atualiza e corrige a ontologia aplicada.

Reforça-se que o encadeamento das atividades não pretende ser o melhor recurso, pois não é este o foco. Elos são uma opção viável e consolidada por estudos anteriores e, como tudo em ciência, são passíveis de melhoramentos. Logo, esses passos não esgotam o tema, pois há atividades paralelas ao ciclo de vida que podem e devem ser realizadas. São elas:

AQUISIÇÃO DE CONHECIMENTO – Adquire o conhecimento sobre o assunto utilizando técnicas de dedução junto aos especialistas de domínio ou por referência à bibliografia relevante. Várias técnicas podem ser utilizadas para a aquisição de conhecimentos, tais como o *brainstorming*, entrevistas, questionários, análise de textos análise e técnicas de indução;

AVALIAÇÃO – Julgamento técnico, baseado em técnicas disponíveis, da qualidade da ontologia;

DOCUMENTAÇÃO – Relata o que foi realizado, como foi feito e o porquê. A documentação associada aos termos representados na ontologia é particularmente importante não apenas para

melhorar sua clareza, mas também para facilitar a manutenção, a utilização e a reutilização.

Ainda que o processo de construção manual de ontologias tenha se desenvolvido e atingido grau de maturidade aceitável nas últimas décadas, essa tarefa constitui árduo e minucioso trabalho que consome recursos financeiros e exige grande dedicação de especialistas de domínio para sua consecução. A automação de tarefas parece ser uma alternativa razoável para minimizar esses problemas mencionados.

APRENDIZADO DE ONTOLOGIAS

De onde vem o termo aprendizado de ontologia? Uma provável origem é de que ele tenha sido inspirado em aprendizado de máquina⁵, que está no campo de análise de dados. Grande parte dos princípios utilizados nessa técnica foi adaptada para a construção e aquisição de ontologias. O fato é que o termo remete à ontologia que foi aprendida ou apreendida.

Assim como na CI, as influências de outras áreas são marcantes na formação do núcleo conceitual do aprendizado de ontologia. Destacam-se as seguintes: a representação do conhecimento, que é objetivo primário de ontologias e mantém estreita relação com a Web semântica; o processamento de linguagem natural (PLN), que é tema central da linguística computacional. Ele é a ferramenta que habilita máquinas a compreender a linguagem natural; e a recuperação da informação, que desenvolve métodos e técnicas que melhoram o desempenho do gerenciamento de motores de buscas e de bases de dados. Esses, popularizados com buscadores de informação na internet, como Google e Yahoo!.

⁵ Do termo em inglês *machine learning*, que é referenciado frequentemente sem tradução.

APRENDIZADO DE ONTOLOGIAS A PARTIR DE TEXTOS

Do texto se vai à ontologia. Conforme discutido, a construção de ontologias requer uma entrada de dados. Neste caso, textos. Eles são igualmente necessários para que se possa aprender termos, conceitos e relações entre eles.

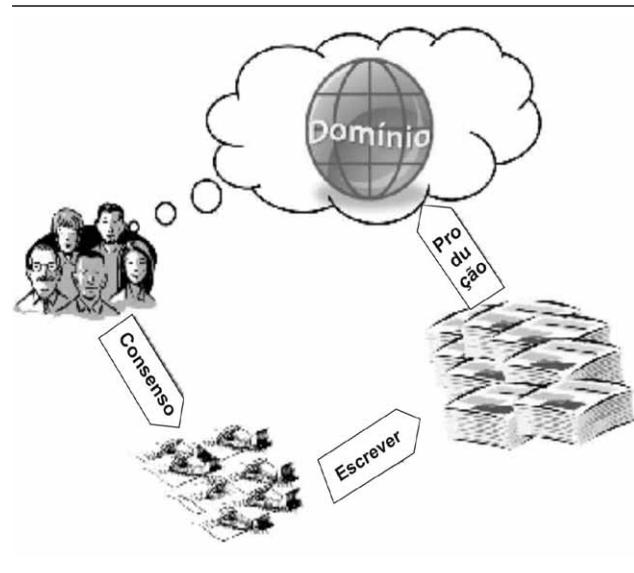
A questão que se suscita é se a entrada de dados é suficientemente representativa de um domínio. Zouaq et al. (2011) afirmam que a qualidade da ontologia gerada depende fortemente da qualidade da fonte de textos.

A figura 2 mostra o fluxo genérico de construção de um domínio a partir de uma entrada de dados textuais.

Refletindo sobre a figura 2, é razoável supor que um grupo de autores que produzem textos de determinado domínio compartilham, em certo grau, a mesma terminologia e conceitos nesse âmbito. Portanto, existe um modelo de domínio implícito no conteúdo resultante dessa produção textual. A tarefa de reconstrução da visão de mundo, ou de domínio, compartilhada por vários autores, pode ser vista como um processo de engenharia reversa (CIMIANO, 2006).

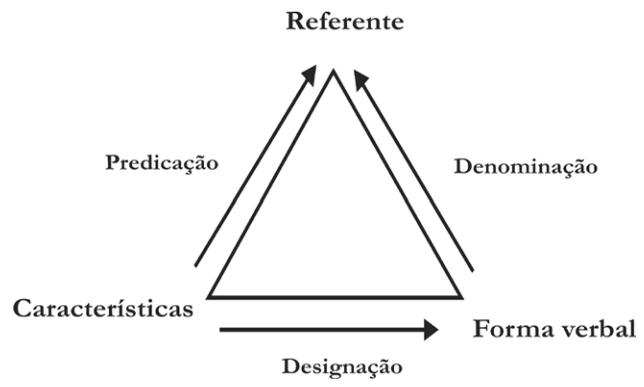
O processo de descoberta de conhecimento apresenta alguns desafios a serem superados. Um deles: o domínio construído a partir de um *corpus* possui o viés dos autores dos textos. Essa questão possui o mesmo problema da construção manual, que estabelece definições segundo o ponto de vista de especialistas de domínio. Isto é, uma visão de mundo limitada ao conhecimento desses profissionais; outro, a ausência de conceitos e relações explícitas no texto. Pois o processo de escrita pode ser visto como atividade de manutenção e explicitação de conhecimento, entretanto apenas parte do conhecimento está efetivamente explícito.

FIGURA 2
Processo de produção textual de um domínio



De acordo com Dahlberg (1978), o homem emprega palavras para traduzir os pensamentos sobre objetos que o circundam. Além disso, utiliza a linguagem para relacionar os objetos aos seus conceitos. Desse modo, formulam-se enunciados sobre atributos necessários ou possíveis dos objetos para que se obtenham as características necessárias dos respectivos conceitos. A figura 3 apresenta o esquema que relaciona o objeto, o signo e o conceito.

FIGURA 3
Triângulo de Dahlberg (DAHLBERG, 1978)



A figura indica que a característica (conceito) é um atributo predicável do referente (o objeto). Ressalta-se que não se trata de uma característica, mas de uma hierarquia de características que estruturam o conceito. No outro vértice do triângulo está a forma verbal (signo), que é a denominação do referente e a designação da característica, de modo a distingui-la de outros objetos.

Nesse sentido, pode-se pensar na escrita como signos da linguagem que expressam a visão de mundo de um grupo de autores de determinado domínio. Esses signos necessitam de interpretação que remete aos conceitos e aos seus correspondentes no mundo, os objetos.

Relação entre conceitos

De acordo com Shamsfard e Barforoush (2003), as relações podem ser estudadas das seguintes maneiras:

Uma relação é um nó em uma ontologia, de modo que seja visto como um conceito e aquela deve ser aprendida como tal.

Uma relação conecta dois ou mais conceitos, de forma que ela deve ser aprendida como um subgrupo de um produto de vários conceitos.

Dahlberg (1978) ensina que, quando se comparam dois conceitos diferentes, há relação entre eles. Nessa comparação, verificam-se nenhuma, uma ou mais características em comum. Logo, existem relações lógicas que auxiliam o estabelecimento dessas relações.

A partir do quadro 1, fica mais claro o estabelecimento de relações semânticas de conceitos. Elas podem ser:

HIERÁRQUICAS – estabelecem o tipo de relação “é-um”. O conceito mais específico herda todas as características daquele mais abrangente. Exemplo: cachorro, mamífero.

PARTITIVAS – existem entre o todo e suas partes. Exemplo: a copa, a aba e o chapéu.

QUADRO 1 Relações lógicas

Identidade	$A(x,x,x)$	$B(x,x,x)$	As características são as mesmas
Implicação	$A(x,x)$	$B(x,x,x)$	O conceito A está contido em B
Intersecção	$A(x,x,o)$	$B(x,o,o)$	Os dois conceitos coincidem algum elemento
Disjunção	$A(x,x,x)$	$B(o,o,o)$	Os conceitos se excluem mutuamente. Nenhuma característica em comum
Negação	$A(x,x,o)$	$B(o,x,o)$	O conceito A inclui uma característica cuja negação se encontra em B

Fonte: (DAHLBERG, 1978).

DE OPOSIÇÃO – dividem-se em contradição – Exemplo: presente e ausente – e contrariedade – Exemplo: claro e escuro.

FUNCIONAIS – aplicam-se aos processos. Exemplo: produto, produtor, comprador e produção.

Uma vez definidos os conceitos e as relações possíveis, torna-se mais clara a tarefa de desenvolvimento de ontologias, que visa estabelecer as definições e comparações entre os conceitos.

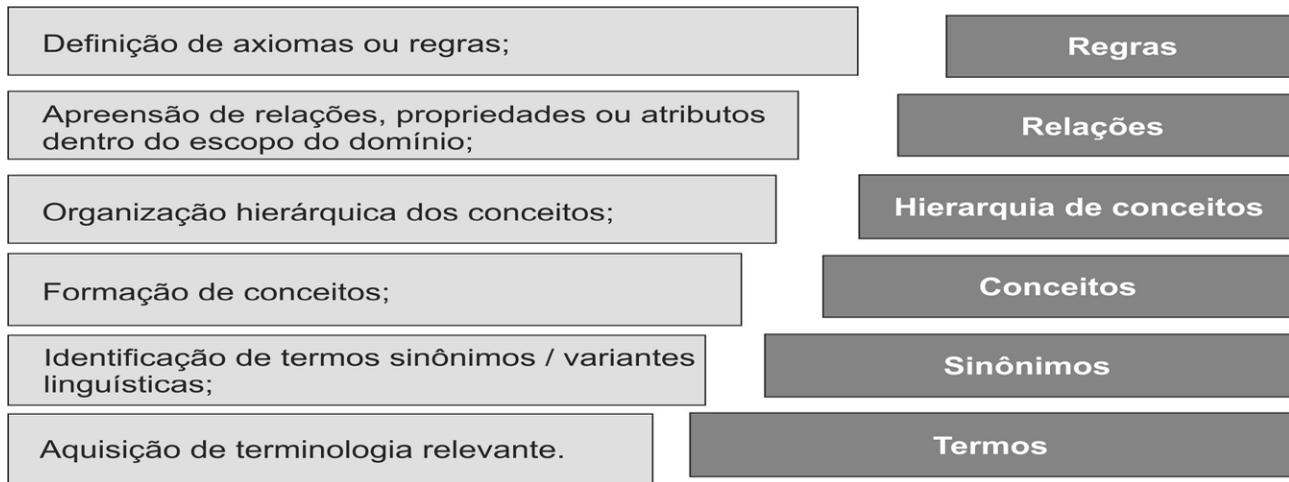
Tarefas na construção de ontologias

Pragmaticamente, há que se estruturar o desenvolvimento de ontologia de modo a possibilitar a automatização das tarefas necessárias para sua construção. O esquema proposto por Buitelaar et al. (2005) objetiva traçar etapas bem definidas, nas quais cada subtarefa venha complementar a anterior, executadas por ordem de complexidade, isto é, a anterior é menos complexa que a atual.

A figura 4, a seguir, mostra a representação dessas tarefas.

De acordo com a figura 4, obtém-se o desenvolvimento completo de uma ontologia por

FIGURA 4

Aprendizado de ontologia em camadas

intermédio de etapas que vão incrementando o grau de complexidade da base para o topo, em relação às respectivas consecuições. Na maioria dos casos, a execução dessas camadas é conceitualmente construída uma sobre a outra, ou seja, uma camada se apoia no resultado obtido na anterior.

Da proposta esquemática da figura 4, infere-se que o núcleo fundamental do desenvolvimento de ontologias está na definição de conceitos e as relações entre eles. Para tanto, existe a necessidade do conhecimento dos signos utilizados para se referir aos conceitos e relações. Isso implica aquisição de conhecimentos linguísticos sobre os termos e eventuais sinônimos que se referem a determinado conceito. Complementarmente, uma ontologia consiste de relações hierárquicas e não hierárquicas. Por último, há fatos não expressos claramente no texto, mas passíveis de dedução ou inferência. Para tal, deve-se definir ou adquirir axiomas ou regras para permitir tais derivações.

Estado da arte

A construção de ontologias pode seguir os passos discutidos. Nesta seção, descreve-se cada uma das camadas e apontam-se as atividades relacionadas que compõem o estado da arte neste campo, em consonância com a visão de Buitelaar et al. (2005).

TERMOS – A extração de termos é pré-requisito para o aprendizado de ontologias a partir do texto. Há muitos exemplos de extração disponível. A maioria é baseada nos métodos de recuperação de informação, velha conhecida dos profissionais da CI para indexação de termos (SALTON; BUCKLEY, 1988). Os termos são as expressões linguísticas que possibilitam a identificação de conceitos e relações e, portanto, são os pilares para construção da base em que se apoia todo o processo de desenvolvimento. As técnicas de processamento de linguagem natural possuem grande relevância para a extração de termos. Isso implica utilização de processamento linguístico na análise de frases que procuram identificar sintagmas nominais que podem expressar termos, estruturas de valências, limitações na ordenação das palavras, características morfossintáticas que buscam evidenciar as estruturas semânticas internas. As ferramentas de PLN atuais apresentam resultados satisfatórios na identificação desses termos. Elas são a combinação de processamentos estatísticos e análises linguísticas. As técnicas mais utilizadas são: *parsers* para extração de termos, *POS tagger* (etiquetadores), algoritmos de identificação de palavras-chave, distribuição estatística de termos e testes estatísticos para identificação de termos compostos;

SINÔNIMOS – A identificação de sinônimos consiste em descobrir termos que denotam o mesmo conceito, ou seja, são considerados sinônimos os termos que compartilham o mesmo significado. A desambiguação é outra meta a ser conseguida, haja vista que o significado de um termo pode variar de um domínio para outro. As principais técnicas utilizadas para descoberta de sinônimos são as de agrupamentos (*clustering*), *Latent Semantic Indexing (LSI)* e a utilização de dicionários especializados de apoio, como tesouros, wordnet, Wikipédia, etc.;

CONCEITOS – Dahlberg (1978) ensina que o conceito é definido como a compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico, nesse caso, termo. De acordo com Cimiano (2006), em aprendizado de ontologias a partir do texto, a indução de conceitos deve observar o seguinte:

– intensão do conceito é a soma total de suas características, que é o conjunto de atributos essenciais que o distinguem no mundo. Por exemplo, cachorro é um animal, vertebrado, mamífero, carnívoro, de quatro patas, da família dos canídeos;

– extensão do conceito é a soma total dos conceitos mais específicos que determinado conceito possui. Ou seja, confirma a intensão do conceito e o torna único. Por exemplo, meu cachorro é um pastor-alemão que se chama Mila;

– conjunto de realizações linguísticas, isto é, de termos para esse conceito.

Na prática, agrupamento de termos pode evidenciar o conceito. Esses grupos se referem a um conceito comum, como é o caso de termos frequentes em textos de matemática, medicina, etc. O uso de dicionários de apoio também é recorrente na identificação de termos que possuem relação semântica.

HIERARQUIA DE CONCEITOS – Taxonomias são utilizadas para organizar o conhecimento

ontológico empregando relações de generalização/especialização por meio das quais o conceito de herança é aplicado. Nesse tipo de hierarquia existe a relação “é-um”, que liga um conceito mais específico a um mais abrangente. Por exemplo, a laranjeira é uma árvore frutífera que é uma árvore. No âmbito da linguística, essas relações são frequentemente referenciadas como:

HIPONÍMIA/HIPERONÍMIA – relação existente entre uma palavra de sentido mais específico e outra de sentido mais genérico, que tem com a primeira traços semânticos comuns (p.ex. mamífero está numa relação de hiponímia com animal).

As principais técnicas se apoiam em análises linguísticas, algoritmos de descoberta de agrupamentos hierárquicos e de coocorrência de termos;

RELAÇÕES – As relações não taxonômicas referem-se a qualquer relação entre conceitos, exceto a relação “é-um”, quais sejam as partitivas, de oposição, funcional e de identidade. As principais técnicas se apoiam em análises linguísticas, algoritmos de descoberta de agrupamentos e de coocorrência de termos;

REGRAS – Regras ou axiomas são utilizados para modelar sentenças que são sempre verdadeiras. Elas podem ser incluídas numa ontologia com o propósito de restringir, de verificar a informação do domínio ou de deduzir uma nova. Nesta camada é possível aprender axiomas no texto. Exemplo: se os termos estão coordenados como na expressão “homem e mulher”, então é provável que eles sejam disjuntos, isto é, sua intersecção é vazia.

CONCLUSÃO

A representação do conhecimento e os assuntos que dela advêm instigam pensadores há vários séculos. A dificuldade de entender os mecanismos cognitivos e sensitivos do ser humano provoca

estranheza e admiração, desde a época dos filósofos gregos até os dias atuais. Muito já foi criado, defendido, refutado, reinventado e ainda não se chegou ao consenso, embora sempre evoluindo no processo de construção do conhecimento.

O desenvolvimento da tecnologia impulsiona pesquisas sobre o tema e disponibiliza máquinas capazes de executar tarefas que seriam impossíveis aos homens. Além disso, diminui o espaço-tempo e coloca comunidades científicas em contato ao redor do mundo. A Internet possibilitou a divulgação da informação de maneira instantânea e acessível a quase todas as pessoas. Nesse sentido, a informação foi gerada em quantidades suficientes para que as máquinas se tornassem parceiras fundamentais na tarefa de apreendê-la.

O estudo de ontologias estabelece importante linha de pesquisa para o incremento da Web Semântica. Seu desenvolvimento e popularização incentivam a busca de soluções que tragam qualidade às estruturas de informações dispersas no mundo digital da Web. As grandes bases de dados, especialmente as bases textuais, são beneficiadas com a evolução de tecnologias que buscam a tradução de textos para uma linguagem que a máquina “compreenda”.

Este tipo de pesquisa ocupa lugar de destaque na CI, cujo domínio é a transmissão do universo do conhecimento humano registrado e concentra-se na manipulação⁶ da informação e não no propósito de conhecer a informação (BATES, 1999).

A combinação de áreas como inteligência artificial, gerenciamento de bases de dados, estatística, recuperação da informação e linguística computacional fundamenta a base teórica para os estudos na construção de ontologias. A automatização de tarefas para a construção ou aprendizado de ontologias vem beneficiar os pesquisadores da área, os quais enfrentam tarefa

complexa que é o desenvolvimento desse campo do saber.

Pelas razões apresentadas em todo o texto, constata-se a aderência do tema com a CI, pois segundo Saracevic (2009), ela é a ciência e prática que lida com o acervo, armazenamento, recuperação e uso efetivo da informação. Ela se concentra no conhecimento e informação registrados e nas tecnologias e serviços relacionados que facilitam seu gerenciamento e uso.

Mais uma vez, as máquinas desempenham papel relevante na realização de tarefas para o desenvolvimento de ontologias. O aprendizado de ontologias a partir do texto extrai termos e definições com base na produção textual de autores de determinado domínio que comungam, em certo grau, a terminologia, conceitos e relações entre eles. Executa-se, então, o caminho inverso da criação manual de ontologias. Pois, basicamente, concentra-se na representação, em forma de signos, do conhecimento tácito de especialistas de domínio. No aprendizado de ontologias, parte-se dos signos para tentar captar o conhecimento que eles explicitaram na forma de texto.

Na última década, o volume de trabalhos sobre o assunto tem crescido, mas ainda há muito a evoluir. O fato é que existem lacunas no método e nas tecnologias para apreensão automática do texto de conceitos e relações entre eles. Outro ponto, a definição e identificação de axiomas e regras que delimitam o domínio, não é assunto trivial. Questionamentos como: Quanto conhecimento está implícito nos textos, de vários autores, que podem realmente ser compartilhados? Ou ainda, as ferramentas disponíveis são capazes de capturar esse conhecimento compartilhado? Ou mesmo, a relação do conhecimento, do texto e do processo de escrita é realmente verdadeira? Essas questões, abertas à discussão, são temas que promoverão vários trabalhos futuros na busca de respostas.

⁶ Representação, organização e recuperação.

REFERÊNCIAS

- BATES, M. J. The invisible substrate of information science. *Journal of the American Society for Information Science*, v. 50, n. 12, p. 1043–1050, 1999.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, v. 284, n. 5, p. 34–43, 2001. ISSN 0036-8733. Disponível em: <<http://www.jeckle.de/files/tblSW.pdf>>. Acesso em: 12/01/2011
- BORST, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. 1997. Tese (Doutorado) – Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands. Disponível em: <<http://doc.utwente.nl/17864/1/t0000004.pdf>>. Acesso em: 05/05/2011
- BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. Ontology Learning from Text: An Overview. In: _____ (Ed.). *Ontology Learning from Text: Methods, Applications and Evaluation*. 3. ed. Amsterdam; New York; Oxford: IOS Press, 2005. Cap. 1, p. 3–12.
- CHAUÍ, M. *Convite à filosofia*. 13. ed. São Paulo: Ed. Ática, 2003. 424 p. I
- CIMIANO, P. *Ontology learning and population from text: algorithms, evaluation and applications*. Karlsruhe, Germany: Springer, 2006. 375 p.
- DACONTA, M. C.; SMITH, K. T.; OBRST, L. J. *The semantic web: a guide to the future of XML, Web services, and knowledge management*. Indianapolis, Indiana: Wiley Pub.-2003. 281 p.
- DAHLBERG, I. Teoria do Conceito. *Ciência da Informação*, v. 7, n. 2, p. 101–107, 1978.
- GRUBER, T. R. What is an ontology? 1993. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 25/01/2011
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *Int. J. Hum.-Comput. Stud.* v. 43, n. 5-/6, p. 625–640, 1995.
- HOUAISS, A. *Dicionário Houaiss da Língua Portuguesa*. Universo On-line, S.D. Internet. Disponível em: <<http://houaiss.uol.com.br>>.
- HOVY, E. Comparing Sets of Semantic Relations in Ontologies. In: GREEN, R.; BEAN, C. A.; MYAENG, S. H. (Ed.). *The semantics of relationships: an interdisciplinary perspective*. Norwell, MA, USA: Springer, 2002. v. 3, cap. 6, p. 91–110.
- MAEDCHE, A.; STAAB, S. Ontology learning for the Semantic Web. *IEEE Intelligent Systems*, IEEE Educational Activities Department, v. 16, n. 2, p. 72–79, 2001.
- NOY, N. F.; HAFNER, C. D. The state of the art in ontology design: a survey and comparative review. *AI Magazine*, p. 53–74, 1997.
- PINTO, H. S.; MARTINS, J. A. P. Ontologies: How can they be built? *Knowledge and Information Systems*, Springer-Verlag New York, Inc., v. 6, n. 4, p. 441–464, 2004.
- SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, v. 24, n. 5, p. 513–523, 1988..
- SARACEVIC, T. Information Science. In: BATES, M. J.; MAACK, M. N. (Ed.). *Encyclopedia of Library and Information Science*. New York: Taylor an Francis, 2009. p. 2570–2586. Disponível em: <<http://comminfo.rutgers.edu/tefko-/SaracevicInformationScienceELIS2009.pdf>>. Acesso em: 21/03/2011
- SHAMSFARD, M.; BARFOROUSH, A. A. The state of the art in ontology learning: a framework for comparison. *The Knowledge Engineering Review*, Cambridge University Press, v. 18, n. 4, p. 293–316, 2003.
- ZOUAQ, A.; GASEVIC, D.; HATALA, M. Towards open ontology learning and filtering. *Information Systems*, v. 36, n. 7, p. 1064–1081, Nov. 2011. ISSN 03064379. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0306437911000391>>.