

O papel das definições na pesquisa em ontologia

Maria Luiza de Almeida Campos

Doutora em Ciência da Informação pelo IBICT/UFRJ . Professora do Departamento de Ciência da Informação da Universidade Federal Fluminense e do Programa de Mestrado em Ciência da Informação da UFF.

Investiga o uso de definição como um fator importante para a expressividade semântica em ontologias. Estudos revelam que as ontologias atuais não possuem padrão para definições. Isso levanta a questão de compatibilidade entre vocabulários que operam em bases cooperativas, como é o caso das pesquisas desenvolvidas em Bioinformática. A literatura tem apresentado alguma discussão em torno das definições em ontologias, mas principalmente como uma análise das definições existentes. Baseada em princípios teóricos, esta pesquisa visa desenvolver um tipo adequado de definição para a construção de ontologias de domínio consistente, ou seja, as definições operacionais.

Palavras-chave: *Ontologia de Domínio; Definição; Compatibilização semântica.*

The role of the definitions in ontology research

Definition is investigated as an important issue for semantic expressiveness in ontologies. Studies reveal that current ontologies do not have standards for definitions. This raises the question of compatibility among vocabularies of systems working cooperatively, as is the case of researches developed in Bioinformatics. Discussions about definitions in ontologies are found in the literature, but mostly on analysis of former definitions. Based on theoretical principles, our research aims at developing a suitable type of definition in the construction of consistent domain ontologies, i.e. operative definitions.

Keywords: *Domain ontologies; Definition; Semantic compatibilization.*

Recebido em 11.10.2009 Aceito 05.03.2010

1 Considerações iniciais

Nos últimos anos, não apenas a oferta de acervos, mas também a de serviços tem proliferado através de portais e sítios na Internet, assim como nas Intranets corporativas. No entanto, a recuperação dos conteúdos e recursos disponibilizados ainda não é realizada de forma satisfatória devido à falta de ferramentas de acesso adequadas que viabilizem maior precisão tanto na especificação da solicitação, quanto na descrição dos conteúdos e recursos disponibilizados. O que vem ocorrendo de modo mais freqüente é a utilização de métodos automáticos de indexação e recuperação de informação, que estão ainda distantes de se constituírem em solução apropriada. Esses métodos se baseiam, principalmente, na freqüência de ocorrência de palavras em documentos, sem contemplar questões ligadas à natureza semântica dos mecanismos de classificação desses recursos. Mesmo quando da utilização de descritores associados aos recursos, a falta de padrões e de um vocabulário comum dentro de cada domínio de negócio torna esses descritores ainda pouco efetivos como elementos de apoio aos mecanismos de recuperação,

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 da difusão de linguagens e ferramentas para a representação e construção de ontologias, as metodologias propostas resultam de pouca utilidade, pois em geral ainda não contemplam diretrizes satisfatórias para a identificação dos conceitos e relacionamentos entre estes, nem tampouco para a criação de definições sistemáticas associadas a esses conceitos. Por conseqüência, as ferramentas têm pouco a contribuir no sentido de orientação do usuário no processo de construção da ontologia, assim como em diretivas para a construção de ontologias de qualidade. (FERNANDEZ; GOMEZ-PEREZ; JURISTO, 1997; DING, 2002a; 2002b; FERNÁNDEZ-LÓPES, 2001; GANGEMI; PISANELLI; STEVE, 1999; GRÜNINGER; FOX, 1995; GUARINO; WELTY, 2002; METAXIOTIS; PSARRAS; ASKNOUNIS, 2001; SURE; STAAB; STUDER, 2002; PASLARU-BONTAS, 2007).

Este artigo visa preencher algumas dessas lacunas, através de estudos que investiguem o uso de definição como um fator importante para a expressividade semântica em ontologias.

No caso das ontologias, as definições propiciam a possibilidade de compatibilização semântica, pois descrevem o conteúdo semântico de um termo. Essa descrição possibilita que agentes inteligentes possam entender o significado de um termo e estabelecer inferências sobre esses significados, pois a definição é composta de características de conceitos, que são também conceitos que se relacionam formando o entendimento semântico dos termos em questão

Desta forma, as definições são de fundamental importância para a elaboração de ontologias consistentes. Entretanto, é fato, que as ontologias existentes ressentem-se, hoje, de um padrão de seu enunciado definatório¹ para a sua elaboração (MICHAEL; MEJINO; ROSSE, 2001; SMITH; KUMAR, 2004; SMITH *et al.*, 2007). Isto é bastante problemático quando se coloca a questão da compatibilização de linguagens que operam em bases cooperativas, como é o caso das pesquisas que vêm sendo desenvolvidas em Bioinformática. Em nossas atividades de pesquisa, estudos vêm sendo desenvolvidos nesse campo (CAMPOS, 2001; 2005; CAMPOS, M. L. M.; CAMPOS, M. L. A.; CAMPOS, L. M., 2005; SALES; CAMPOS; GOMES, 2006; CAMPOS; GOMES, 2007; CAMPOS *et al.*, 2009). Estas nos levam a constatar que, para se trabalhar em domínios específicos, será necessário desenvolver padrões para enunciados definatórios que venham a atender à especificidade de cada domínio. Para pensar em um padrão de mais alto nível, temos por hipótese que é necessário explicitar um número mínimo de elementos para a descrição de um objeto, tais como: características que indicam o gênero próximo e a diferença específica do conceito em análise, seus componentes (caso o conceito seja um objeto concreto ou abstrato) ou suas etapas (caso seja um processo ou atividade) e, por fim, a finalidade de aplicação no contexto no qual será aplicado.

Este artigo tem por objetivo apresentar estudos teóricos no que tange à problemática definatória em ontologias formais de domínio, fruto das pesquisas no âmbito das ontologias de domínio genômico, apoiadas pelo CNPq².

2 Ontologia formal e de domínio

O termo ontologia possui uma conceituação própria na Ciência da Computação, especificamente na Inteligência Artificial, diferentemente da Filosofia. Em Inteligência Artificial "uma ontologia é uma rede de definições de um vocabulário que expressa um consenso da comunidade

¹ Entendemos aqui um enunciado definatório, como o conjunto de características que descrevem um conceito em um dado contexto. Tem por função além de permitir o posicionamento do conceito em um sistema de conceitos, o de possibilitar o entendimento de toda a complexidade daquele conceito em um dado contexto, ou seja, que elementos ônticos (relações funcionais e partitivas) possam ser agregados aos elementos lógicos (relações de gênero-espécie) (DAHLBERG, 1983).

² Projetos de Pesquisa: Período 2005-2008 - Integração de Ontologias: o domínio da bioinformática e a problemática da compatibilização terminológica, e 2009-2011 - Ontologia e o papel das definições: a descoberta do conhecimento no domínio de Tripanosomatídeos como apoio às pesquisas em Bioinformática no Brasil.

sobre o domínio de conhecimento". (WEINSTEIN, 1998, p. 256). Uma ontologia é, assim, um conjunto de conceitos padronizados, termos e definições aceitas por uma comunidade em particular. A mais freqüente definição de ontologia é a de Gruber (GRUBER, 1993, p. 199) "uma ontologia é uma especificação de uma conceituação".

Uma conceituação é uma abstração, uma visão simplificada do mundo que se representa para satisfazer a um ou mais dos seguintes propósitos: "permitir que múltiplos agentes compartilhem seu conhecimento; ajudar as pessoas a compreender melhor certa área de conhecimento; ajudar as pessoas a atingir um consenso no seu entendimento sobre uma área de conhecimento" (SMITH, 1996 *apud* FALBO, 1998, p. 47). Em Lógica, uma conceituação identifica o objeto e as relações que existem no universo lógico (WEINSTEIN, 1998).

A literatura, muitas vezes, vem denominando ontologias como 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. Esses tipos de ontologias são chamadas, também, de ontologias informais (WEINSTEIN, 1998) ou ontologias lingüísticas (GUARINO, 1998b). Na verdade, o termo ontologia vem sendo usado para caracterizar qualquer classificação de um domínio ou uma lista de termos definidos.

No campo da Inteligência Artificial, o que se denomina por ontologia é considerado ontologia formal, ou seja, aquela que define vocabulário com lógica. A sintaxe exata e a semântica dependem da representação da linguagem, 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. 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.

Além disso, apesar dos conceitos e das relações formarem a base de uma ontologia, uma característica essencial das ontologias é a definição de axiomas.

Simplemente propor uma taxionomia ou um conjunto de termos básico, não constitui uma ontologia. Axiomas devem ser providos para definir a semântica dos termos. Os axiomas especificam definições de termos na ontologia e restrições sobre sua interpretação (OLIVEIRA, 1999).

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.

As ontologias podem ser classificadas, segundo a abrangência do domínio que se quer representar, em: Ontologia de Alto Nível, Ontologia de Domínio/Tarefa e Ontologia de Aplicação. As Ontologias de Alto Nível descrevem conceitos muito gerais como espaço, tempo, matéria, objeto, evento, ação, etc., que são independentes de um problema particular ou domínio. As Ontologias de Domínio e Ontologias de Tarefas descrevem, respectivamente, o vocabulário relacionado com um domínio genérico (como medicina ou automóvel) ou com uma tarefa, atividade genérica (como diagnóstico ou venda) pela especialização dos termos introduzidos na ontologia de alto nível. As Ontologias de Aplicação descrevem conceitos que dependem tanto de um domínio particular como de uma tarefa, que são muitas vezes especializações de ambas as ontologias relacionadas (GUARINO, 1997).

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

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) e sobre as categorias de meta-nível para modelar o mundo (conceitos, propriedades, qualidades, estados, papéis e partes) (GUARINO, 1998a).

3 Representação do conhecimento

Podemos afirmar que um corpo de conhecimento representado é baseado em uma conceituação dos objetos, dos conceitos e de outras entidades, que se supõe existir em alguma área de interesse e dos relacionamentos que os ligam. Desta forma, toda base de conhecimento é comprometida com alguma conceituação. Esta deve poder ser representada para atender às necessidades de sistemas automatizados.

Nesses sistemas, são os mecanismos de representação de conhecimento que podem tornar os processos de formalização, sobre os objetos e suas relações, em contextos pré-definidos, facilmente representados. No âmbito das ontologias esses mecanismos permitem a sistematização dos conceitos e, conseqüentemente, a elaboração de definições consistentes, visando possibilitar futuras inferências sobre o domínio.

Segundo DAVIS *et al.* (1992, p. 182, tradução nossa), o conceito de representação de conhecimento pode ser mais bem entendido a partir das seguintes definições ligadas aos papéis que poderá desempenhar:

1. Uma representação de conhecimento é um mecanismo usado para se raciocinar sobre o mundo ao invés de agir diretamente sobre ele. Neste sentido, ela é, fundamentalmente, um substituto para aquilo que representa. Este papel conduz, naturalmente, a dois importantes aspectos. O primeiro diz respeito a sua identidade projetada. Deve haver alguma forma de correspondência especificada entre o substituto e

seu referente planejado no mundo. Esta correspondência é a semântica da representação. O segundo é a fidelidade. Fidelidade perfeita é, em geral, impossível, tanto na prática quanto em princípio. A única representação completamente precisa de um objeto é o objeto em si. Qualquer outra representação é imprecisa e, inevitavelmente, contém simplificações;

2. Uma representação de conhecimento é uma resposta à pergunta "Em que termos devo pensar sobre o mundo? Isto é, um conjunto de compromissos ontológicos. Uma vez que toda representação é uma aproximação imperfeita da realidade, ao selecionar uma representação, estamos tomando um conjunto de decisões sobre como e o que ver no mundo. Ou seja, selecionar uma representação significa fazer um conjunto de compromissos ontológicos. Esses compromissos determinam o que pode ser visto, enfocando alguma parte do mundo em detrimento de outras. Esta forma de ver o mundo não é apenas um efeito colateral da escolha de uma representação; ao contrário, o efeito focalizador é a parte essencial do que a representação oferece, já que a complexidade do mundo real é esmagadora. Assim, o comprometimento ontológico feito por uma representação pode ser uma de suas mais importantes contribuições.
3. Uma representação de conhecimento é uma teoria fragmentada de raciocínio que especifica que inferências são válidas e quais são recomendadas. Uma representação é motivada por alguma percepção de como as pessoas argumentam ou por alguma crença sobre o que significa raciocinar de forma inteligente. A teoria de raciocínio inteligente embutida em uma representação é geralmente implícita, mas pode se tornar mais evidente pelo exame de três componentes: a concepção de inferência inteligente, o conjunto de inferências que a representação sanciona e o conjunto de inferências que ela recomenda. Enquanto as inferências sancionadas indicam o que pode ser inferido, as inferências recomendadas dizem respeito ao que deve ser inferido. Esta orientação é necessária, pois o conjunto de inferências sancionadas é tipicamente muito grande para ser usado indiscriminadamente. Estes componentes podem ser vistos, também, como as respostas da representação a três questões fundamentais: (i) O que significa raciocinar de forma inteligente? (ii) O que podemos inferir a partir do que conhecemos? e (iii) O que devemos inferir a partir do que conhecemos?
4. Uma representação de conhecimento é um meio de computação pragmaticamente eficiente. Na realidade, esta questão aborda a utilidade prática da representação. Se ela torna coisas possíveis mas não facilmente computáveis, então,

a representação pode não ser de muita valia para o problema em mãos.

5. Uma representação de conhecimento é um meio de expressão, i.e., uma linguagem na qual pode-se dizer coisas sobre o mundo. Esta é, também, uma questão de utilidade prática da representação. Se ela permite expressar certas situações, mas isto não é feito facilmente, então, como usuários, muitas vezes não podemos saber se a representação não é capaz de expressar alguma coisa que gostaríamos de dizer ou simplesmente não sabemos como usá-la. Uma representação é a linguagem na qual nos comunicamos e, assim, devemos ser capazes de falar sem esforço heróico.

Por outro lado, a representação do conhecimento pode ser classificada de acordo com os tipos de primitivas oferecidas ao usuário, segundo Brachman (1979), em quatro níveis: lógico, epistemológico, ontológico e conceitual.

O nível lógico é o nível da formalização. Entretanto, não existe preocupação com a semântica em termos dos conceitos e de suas relações, uma vez que todo o foco está centrado em uma dada "sintaxe" que possibilite uma verdadeira ação do pensar. No nível epistemológico, a noção genérica de um conceito é introduzida como uma primitiva de estruturação de conhecimento, ou seja, ele é o nível da estruturação. O nível ontológico tem por objetivo restringir o número de possibilidades de interpretação do conceito dentro de um dado contexto, a partir de um formalismo que pretende representar o conteúdo do conceito. No nível conceitual, independentemente de um formalismo, os conceitos possuem, *a priori*, uma interpretação definida. O esqueleto da estrutura dos conceitos em um determinado domínio está dado e o conhecimento é expresso na forma de uma especificação desse esqueleto.

Na perspectiva tanto da Ciência da Informação quanto da Ciência da Computação, interessam-nos aqueles níveis nos quais uma estrutura de conhecimento possa ser sistematizada e representada a partir de contextos específicos. Segundo Newel (1982), os níveis epistemológico e ontológico são aqueles que permitem a representação de conhecimento estruturada e formalizada.

No nível epistemológico, como vimos, especifica-se a estrutura dos conceitos e seus inter-relacionamentos. No nível ontológico, avança-se um pouco mais no processo de organização e classificação de um determinado domínio e acrescenta-se a definição dos conceitos que nele estão inseridos. Enquanto o nível epistemológico é o de estruturação, o nível ontológico é o de significação.

Outro aspecto que deve ser observado sobre a natureza do conhecimento diz respeito ao modo como o conhecimento está reunido em um dado contexto. Em alguns contextos, o núcleo temático existente é formado a partir de um corte paradigmático, ou seja, o conhecimento se apresenta como uma área de saber já estabelecida, como uma área denominada de área canônica, como a Ciência da Informação ou a Ciência

da Computação, por exemplo. Em outros contextos, o núcleo temático existente é formado a partir de um corte sintagmático do universo de conhecimento. Isto pode ser explicado como a reunião de saberes de diversas áreas canônicas em que, o que está em jogo é um dado objetivo, uma missão ou problema, uma questão ou um contexto de atividades que se quer representar, como por exemplo: o domínio dos Tripanosomatídeos.

O modo como o núcleo temático de uma ontologia é constituído é de importância capital para a sua elaboração. Essa organização influenciará diretamente as relações que se estabelecem entre os conceitos e, conseqüentemente, nas definições desses conceitos e em sua estruturação no domínio.

4 Ontologia e definição

Como apresentado em diversos estudos, ontologia (CORAZZON, 2000; GRUBER, 1993; GUARINO, 1997; 1998a; SWARTOUT; TATE, 1999; VICKERY, 1997) 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, através da definição de um conjunto de termos que pudessem representar domínios e tarefas a serem executadas por esses programas.

Atualmente, as ontologias vêm sendo utilizadas em diversas áreas. Guarino (1998), em seu trabalho, afirma que esta ontologia com "o" minúsculo, diferentemente da ontologia com "O" maiúsculo, campo de estudos da Filosofia, 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. Ao longo dessa trajetória, a Ciência da Informação desenvolveu um arcabouço teórico e metodológico consubstanciado em obras como as de Otlet (1934), Ranganathan (1967), Dahlberg (1978), Hjørland (2002) e do *Classification Research Group* (CRG) (WILSON, 1972), entre outros, que potencialmente têm grande aplicabilidade no desenvolvimento de ontologias.

Como formalismo de representação do conhecimento, uma ontologia é um tipo de sistema baseado em conhecimento. A ontologia...

[...] é o coração de qualquer descrição do conhecimento: o conhecimento está intimamente relacionado com a ontologia, uma vez que é necessariamente expresso em termos desta ontologia. Portanto, conceber a ontologia de um domínio é um ponto chave no processo de aquisição de conhecimento, durante a etapa de extração do conhecimento (BOUAUD *et al.*, 1994).

Bouaud *et al.* (1994) discutem princípios para a aquisição e estruturação de ontologias que, em certa medida, abordam a importância de definições em ontologias de domínio a partir de questões que problematizam como é possível representar conhecimento para que um mecanismo automatizado possa “compreender” significados. Seus argumentos remontam à própria Inteligência Artificial.

De um modo geral, a Inteligência Artificial procura soluções usando o conhecimento que, em geral, é expresso em linguagem natural. Como não existe qualquer expressão operacional desse conhecimento, isto é, o domínio não é formal, temos que lidar com a riqueza semântica da língua e suas múltiplas interpretações potenciais. O problema básico é que o conhecimento humano para o qual queremos prover um programa não é formal. O conhecimento que as pessoas têm sobre os objetos em um domínio é principalmente *descritivo*. As pessoas raciocinam com este conhecimento descritivo empregando as leis semânticas do sentido como analogia, metáfora, inferência prototípica, etc. Em domínio não formal, não temos definições completas para as categorias de objetos considerados no conhecimento descritivo. Para que um computador lide com essa espécie de conhecimento, precisamos padronizá-lo, precisamos estar de acordo com as noções que ele mobiliza e com seu uso. O conhecimento precisa ser considerado como uma noção objetiva, isto é, compartilhada. Basicamente, temos que fixar o significado dos termos para que todos tenham a mesma compreensão quando usarem um termo. O conhecimento será descrito adequadamente se seu objetivo e seus aspectos forem explicitados. Padronizar é apenas isso. O uso de termos, conseqüentemente, não se apoiará em todo o potencial da interpretação semântica, mas na manipulação de suas descrições explícitas. No entanto, as restrições postas ao conhecimento pelo processo de padronização não devem modificar o significado das noções representadas (BOUAUD *et al.*, 1994).

Outros trabalhos, que ressaltam a importância do estabelecimento de definições coerentes para a elaboração de ontologias, vem sendo apresentados na literatura (MICHEL *et al.*, 2001; KOHLER *et al.*, 2006; HOVY, 2002; SMITH; ROSSE, 2004). Nesses trabalhos, questões relacionadas à importância da descrição dos aspectos (características,

propriedades) dos objetos vêm sendo discutidas visando permitir o raciocínio automático, fundamental para que as ontologias possam possibilitar inferência e descoberta de conhecimento. Os estudos mostram que as definições, em ontologias, devem funcionar como um sistema integrado, no qual, cada elemento da definição deve estar relacionado como em um sistema.

Michael, Mejino e Rosse (2001) em seu artigo "*The role of definitions in Biomedical concept representation*" registram a diferença entre a produção de definições para atender às necessidades de dicionários e léxicos em contraposição àquelas que pretendem atender às ontologias. Argumentam que, nos dicionários, a unidade de informação é uma palavra e que, neste sentido, as definições especificam os significados das palavras. As palavras são apresentadas de forma alfabética e as definições não possuem nenhuma relação com as palavras vizinhas. Essa forma de definir atende às necessidades dos dicionários, nos quais a definição é realizada de forma isolada, sem qualquer relação com as outras palavras. Desta forma, em dicionários são apresentados diferentes significados para uma palavra. Em ontologias, a unidade de informação é o conceito, cuja seqüência de conceitos representa um tipo de hierarquia e essa hierarquia é partilhada por grupos de conceitos. A consistência dessa hierarquia depende da definição explícita das propriedades (atributos, características) que descrevem os conceitos, fornecendo a base sobre a qual eles podem ser agrupados em conjunto, diferenciando-se um conceito do outro. O papel das definições em uma ontologia é, portanto, precisar esses atributos de forma coerente assegurando, assim, principalmente uma herança através de um tipo de hierarquia. A coerência nas definições e, por conseguinte, na classificação, exige uma perspectiva unificadora requerida pelo contexto a ser especificado para a representação dos conceitos. Assim, ao contrário dos dicionários, uma definição de um conceito dentro de uma ontologia pode ser considerada incompleta se todos os pais (conceitos superordenados) não tiverem sido também definidos. Esse aspecto é um dos fatores que permite a consistência da estrutura lógica, muito importante para se possibilitar o raciocínio automático.

Em ontologias, diferentemente de outras linguagens que também têm 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 em uma ontologia. Assim, a definição, em ontologias, necessita ser formalizada para permitir a interpretação, também, por parte de agentes inteligentes e não somente por humanos. Ela necessita de um padrão para seus enunciados definitórios e de uma linguagem lógica que possa apoiar o entendimento da definição, pelos agentes inteligentes.

No caso das ontologias, as definições propiciam a possibilidade de entendimento semântico, pois representam o conteúdo semântico de um termo. Essa representação possibilita que agentes inteligentes possam entender o significado de um termo e estabelecer inferências sobre esses

significados, pois a definição é composta de características de conceitos, que são também conceitos que se relacionam formando o entendimento semântico dos termos em questão.

4.1 Definições Conceituais e sua importância na pesquisa em ontologia

Até a década de 1960, as definições tinham um caráter mais conceitual e filosófico e evidenciavam o que pensar acerca de um conceito. Entretanto, para atender às necessidades da ciência foi necessário um esforço teórico visando à elaboração de definições operacionais – aquelas cujos elementos da definição não identificam somente o gênero próximo e a diferença específica – mas, também, elementos que procuravam definir certas operações nas quais o conceito seria aplicado (HEGENBERG, 1974). Os problemas da definição tem sido objeto de estudos diversos na língua geral. Na língua especializada, nos últimos 40 anos, a definição passou a ser objeto de estudo sistemático. Alguns estudos seminais, aqui apresentados relacionando os problemas definitórios e suas características em um dado domínio, pareceram-nos bastante apropriados, tendo em vista os objetivos desta pesquisa.

Em 1982, o *Groupe Interdisciplinaire de Recherche Scientifique et Appliquée en Terminologie* (GIRSTERM), no Canadá, realizou um Colóquio Internacional de Terminologia sobre o tema "Problemas da definição e da sinonímia em terminologia" com o objetivo de discutir as especificidades da definição em terminologia. Por exemplo, o que definir? Como definir? Por que definir? A sinonímia dos termos se compara à sinonímia das palavras? Essas questões se apresentam no cotidiano dos terminólogos no exercício de seu trabalho (CAMPOS, 2001).

Nesse Colóquio, Dahlberg (1983) apresenta uma longa exposição sobre o caráter e os requisitos da definição terminológica. E faz isso à luz da Teoria do Conceito (DAHLBERG, 1978). Segundo essa Teoria, a definição terminológica seria a definição analítica/conceitual, pois ela incorpora os três elementos do conceito - o referente, as características e o termo - ao contrário da definição nominal, que contempla apenas o termo e uma equivalência textual, e da definição ostensiva, que incorpora apenas o referente e o termo.

Definição conceitual (ou definição real) = df definição na qual o *definiens* contém as características de um referente nomeado pelo *definiendum* (DAHLBERG, 1983, p. 16).

A partir dessa explicação, Dahlberg apresenta três tipos de definições: definição nominal; definição ostensiva; definição conceitual. A definição nominal é aquela em que o *definiendum* é uma expressão verbal

e o *definiens* é uma equivalência textual deste termo, como por exemplo, $A = B$. A definição ostensiva é aquela cujo *definiens* é estabelecido apontando-se para o referente nomeado pelo *definiendum*, ou seja, $C = A$. Já a definição conceitual, também denominada de definição real, ocorre quando o *definiens* contém as características necessárias de um referente nomeado pelo *definiendum*, ou seja, $C = B$ de A .

No caso das questões que envolvem ontologias, interessa-nos o estudo das definições conceituais e das nominais, pois essas permitem a compatibilização no plano semântico e no plano lingüístico, respectivamente.

Dahlberg (1981) afirma que, para se definir conceitos, é necessário o conhecimento de: a) os referentes dos conceitos; b) a(s) categoria à(s) qual (is) pertence um conceito; e c) a expressão verbal adequada a um certo grupo de usuários. E esclarece:

a fim de acumular conhecimento sobre o referente de um conceito, deve-se coletar enunciados sobre este referente A , de modo a identificar suas características. A informação sobre B , a(s) categoria(s), ajuda(m) a identificar o gênero próximo ao qual o conceito deve se relacionar em última instância. O conhecimento de C determina a expressão verbal apropriada a ser usada (DAHLBERG, 1981, p.16).

Defende a autora, portanto, que os conceitos se relacionam em um sistema de conceitos por suas características, que são também conceitos.

Segundo a Teoria do Conceito, as características relevantes do conceito são, então, os elementos constitutivos da definição. O ponto principal no estabelecimento das definições dos conceitos está, portanto, na identificação das características. Ela fornece um padrão para definição, classificando-as conforme a categoria do conceito: definição genérica, partitiva, funcional (DAHLBERG, 1983). A definição genérica permite identificar a categoria do conceito, a partitiva, os componentes do conceito definido, e a funcional insere o conceito como elemento integrador no contexto analisado, ou seja, ela permite que se identifique, na definição, a função/finalidade do conceito, dentro da área em questão. Com isso, a definição terminológica forneceria a base para se estabelecer os sistemas terminológicos. Tanto para Dahlberg (1981), como para Natanson (1983) e Drozd (1983), que participaram do Colóquio anteriormente citado, a definição terminológica, por excelência, é a definição analítica (GOMES; CAMPOS, 1990)

Dahlberg, na verdade, é integrante de um grupo de estudiosos que, já há algum tempo, vinham dando à definição um objetivo mais prático, relacionado com as demandas da terminologia e à necessidade de tornar claras as definições técnicas e científicas e não mais um objetivo de interesse estritamente filosófico. No campo da Terminologia, temos o trabalho pioneiro de Wüster (1981) e, no campo da própria Ciência, temos o movimento chamado operacionismo que, segundo Hegenberg (1974),

surge a partir de 1960, fruto direto do positivismo lógico defendido pelo Círculo de Viena, parente próximo do pragmatismo e do instrumentalismo, como apresentado anteriormente.

O operacionismo representa uma formulação recente de certos traços essenciais do método experimental e do empirismo de modo geral, caminhando na direção do pragmatismo e do instrumentalismo (Peirce, James e Dewey). A forma que Bridgman dá ao critério empírico de significado, embora original, tem muitos pontos de contato com o que Peirce afirma em seu 'How to make our ideas clear' (HEGENBERG, 1974, p.104).

Ainda segundo Hegenberg (1974), as definições procedentes desse movimento denominam-se "operativas".

A idéia fundamental que norteia o operacionismo é simples, a saber, a de que não conhecemos os significados de um conceito a menos que possamos especificar as operações que foram realizadas, por nós ou por outrem, ao aplicar o conceito a qualquer situação concreta (HEGENBERG, 1974, p.105).

Desta forma, as definições operativas revelam a maneira de aplicação dos conceitos, dentro de um dado contexto (uma área de assunto), ou como se dá o uso de tais conceitos o que, mais tarde, Dahlberg denominará de relação funcional.

Gorkova (1980) apresenta ainda duas abordagens para as definições conceituais: a epistemológica e a pragmática. A abordagem epistemológica seria adequada aos conceitos básicos, gerais, da ciência. Para os conceitos específicos, bastaria uma definição pragmática, ou uma "explicação adequada". Kandelaki (1970) desenvolve esse princípio, que parece nortear a Escola Russa de Terminologia.

De fato, a definição epistemológica e a definição pragmática são complementares para as terminologias de uma área técnica. As definições genérica e partitiva descrevem o conteúdo de um conceito. Porém quanto à aplicabilidade do conceito dentro da área em que está definido, é a definição funcional/operativa que permite desempenhar aquele papel.

Observa-se, então, que os diferentes modos de definir podem levar aos diferentes tipos de definição. Não é possível privilegiar-se somente um tipo de definição, porque os conceitos de uma área de assunto são de níveis diversos (conceitos gerais e específicos) e de natureza categorial diversa (objeto, propriedade, processo).

É preciso, portanto, chegar a um modelo de enunciado capaz de atender às especificidades da área. Para tanto, consideramos que a forma definitória deve ser ditada pelo próprio nível do conceito, como, principalmente, por sua natureza categorial, para que se obtenham os

elementos necessários à caracterização do conceito e, quando pertinente, sua função, numa dada área de assunto. Em cada caso, deve-se tentar identificar os modelos mais apropriados de definição, ou seja, os padrões de enunciados definitórios em cada contexto.

4.2 Formação de enunciados definitórios

Um enunciado definitório pode ser explicitado como o conjunto de características que descrevem um conceito em um dado contexto. Sua função não é de somente permitir o posicionamento do conceito em um sistema de conceitos como as definições conceituais se propõem, mas de ampliar toda a complexidade de entendimento daquele conceito em um dado contexto, ou seja, quais elementos ontológicos podem ser agregados aos elementos lógicos (relações de gênero-espécie). Em Ontologias Formais, a elaboração de enunciados definitórios para cada contexto de aplicação se faz necessária devido à sua função de representação da realidade contextual.

Para a elaboração de um enunciado definitório, é necessário que se reúnam os elementos constitutivos que explicitam as características de um conceito, nos quais algumas características possibilitam identificar o posicionamento do conceito em um sistema de conceitos, e outras características possibilitam ter uma visão mais pragmática do uso do conceito, dentro de uma dada área. Teremos assim, reunidas em um mesmo enunciado definitório, definições do tipo analíticas/conceituais e operativas, tendo-se as primeiras a função de posicionamento do conceito e a segunda a função do uso que é atribuído ao conceito em questão. Esta última característica, ou seja, a de uso/aplicação, é de suma importância para as ontologias formais de domínios.

5 Considerações Finais

As pesquisas até agora realizadas apontaram para a necessidade de estudos mais aprofundados no que tange às definições, pois são elas, além das relações conceituais em ontologias, que evidenciam o conteúdo semântico de um conceito. A literatura tem apresentado alguma discussão em torno das definições em ontologias, mas principalmente como uma análise das definições existentes. Em nossos estudos, até agora podemos apontar a importância dos estudos definitórios para a elaboração de ontologias de domínios consistentes, além de avançar, apoiado por estudos teóricos, determinando uma forma mais adequada de tipo de definição para as ontologias de domínio: as definições operativas.

Pretende-se observar, no desenvolvimento dos estudos, quais elementos deverão constituir um enunciado definitório que possam revelar, não apenas o posicionamento do conceito no sistema, mas, também, a descrição de fatos que envolvem tal conceito em um dado contexto, aplicando-se os princípios definitórios conceituais e operativos

em um dado domínio. Consideramos que essa perspectiva poderia recuperar o significado original do termo ontologia, ou seja, a representação do “ser no mundo”.

6 Referências

BOUAUD, J. *et al.* Acquisition and Structuring of an ontology within conceptual graphs. In: WORKSHOP ON KNOWLEDGE ACQUISITION USING CONCEPTUAL GRAPH THEORY, 1994. *Proceedings of ICCS'94...* College Park: MD, University of Maryland, out.1994. p. 1-25.

BRACHMAN, R. J. *On the epistemological status of semantic networks.* In: FINDLER, N. V. (Ed.) *Associative networks: representation and use of knowledge by computers.* New York : Academic Press, 1979. p. 3-50.

CAMPOS, M. L. A. *Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização da autoria.* 2001. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.

CAMPOS, M. L. A. A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 6., 2005, Florianópolis. *A política científica e os desafios da sociedade do conhecimento.* Florianópolis: ANCIB, 2005.

CAMPOS, M. L. M.; CAMPOS, M. L. A.; CAMPOS, L. M. Web semântica e a gestão de conteúdos informacionais. In: MARCONDES, C. H. *et al.* (Orgs.). *Bibliotecas digitais: saberes e práticas.* Salvador, BA; Brasília: EDUFBA; IBICT, 2005. p. 55-75.

CAMPOS, M. L. M. *et al.* Aspectos metodológicos no reuso de ontologias: um estudo a partir das anotações genômicas no domínio dos tripanosomatídeos. *RECIIS - Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, v. 3, p. 64-75, 2009.

CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 8., 2007, Salvador. *Anais...* Salvador: ANCIB, 2007.

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, 1978.

_____. Conceptual definitions for interconcept. *International Classification*, v. 8, n. 1, p. 16-22, 1981.

_____. Terminological definitions: characteristics and demands. In: PROBLÈMES de la définition et de la synonymie en terminologie. Québec: GIRSTERM, 1983. p. 13-51.

DAVIS, H. *et al.* Towards an integrated environment with open hypermedia systems. In: ACM CONFERENCE ON HYPERTEXT, 1992. *Proceedings of...* Milan, Italy, dec. 1992.

DING, Y.; FOO, S. Ontology research and development. Part 1 - a review of ontology mapping and evolving. *Journal of Information Science*, v. 28, n. 2, p. 123-136, 2002a.

_____. Ontology research and development. Part 2 - a review of ontology generation. *Journal of Information Science*, v.28, n.5, p. 375-388, 2002b.

DROZD, L. Terminological synonyms and the function of definition: theses. In: COLLOQUE INTERNATIONAL DE TERMINOLOGIE, Québec, 23-27 maio 1982. *Actes...* 1983. p. 87-100.

FALBO, R. A. *Integração de conhecimento em um ambiente de desenvolvimento de software*. 81 f. 1998. Tese (Doutorado em Ciências) - COPPE/UFRJ, Rio de Janeiro, 1998.

FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. In: SYMPOSIUM ON ONTOLOGICAL ENGINEERING, 1997. *Proceedings of the AAAI1997 Spring...* 1997. p. 33-40.

FERNÁNDEZ-LÓPES, M. Overview of methodologies for building ontologies. In: WORKSHOP ON ONTOLOGIES AND PROBLEM-SOLVING METHODS: Lessons Learned and Future Trends. *Intelligent Systems*, 2001. p. 26-34.

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.

GOMES, H. E.; CAMPOS, M. L. A. Classificação e terminologia. In: SIMPÓSIO DE RITERM, 2., Brasília, 1990. *Anais...* Brasília: [s.n.] , 1990.

GORKOVA, V. I. *Some methodological recommendations*. In: FID. Essential problems in terminology for informatics and documentation. *Moscow: [s.n.]*, 1980. p.7-22.

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. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, 1995. *Proceedings...* Montreal: [s. n.], 1995.

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*. Berlin: Springer Verlag,1997. p. 139-170.

GUARINO, N. Formal ontology and information systems. In: FOIS '98, 1., 1998, Trento, Italy. *Proceedings...* Amsterdam: IOS Press; Tokyo: Omsha, 1998a . p. 3-15.

GUARINO, N. Some ontological principles for designing upper level lexical resorces. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUTION, 1., 1998, Granada. *Proceedings...*1998b.

GUARINO, N.; WELTY, C. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, v. 45, n. 2, p. 61-65, 2002.

HEGENBERG, L. *Definições: termos teóricos e significado*. São Paulo: Cultrix, 1974.

HJØRLAND, B. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p. 422– 462, 2002.

HOVY, E. Comparing sets of semantic relations in ontologies. In: GREEN, R. *Interdisciplinary perspective*. Kluwer: [s.n.], 2002. p. 91-110.

KANDELAKI, T. L. Les sens des termes et les systèmes de sens des terminologies scientifiques et techniques. In: RONDEAU, G.; FELBER, H. *Textes choisis de terminologie: I Fondements théoriques de la terminologie*. Québec: GIRSTERM, 1981. p. 133-184.

KOHLER, J. *et al.* Quality control for terms and definitions in ontologies and taxonomies. *BMC Bioinformatics*, v. 7, n. 212, p.1-12, 2006.

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.

MICHAEL, J.; MEJINO, J. L. V.; ROSSE, C. The role of definitions in biomedical concept representation. *JAMIA Symposium Supplemen*, v. 01, p. 463-467, 2001.

NATANSON, M. E. Rapports entre la définition logique et la délimitation des concepts. In: COLLOQUE INTERNATIONAL DE TERMINOLOGIE, 1982, Québec. *Problèmes de la définition et de la synonymie en terminologie. Actes...* Université Laval, Québec, 23-27 mai 1982. p. 55-65.

NEWELL, A. The knowledge level. *Artificial Intelligence*, v. 18, p. 87-127, 1982.

OLIVEIRA, K. *Modelo para a construção de ambientes de desenvolvimento de software orientados a domínios*. 1999. Tese (Doutorado em Sistemas e Computação) - UFRJ/COPPE Sistemas, out. 1999.

OTLET, P. *Traité de documentation: le livre sur le livre, theorie et pratique*._Bruxelles: Editiones Mundaneum, 1934.

PASLARU-BONTAS, E. *A contextual approach to ontology reuse: methodology, methods and tools for the semantic web*. 2007. Dissertation (Fachbereich Mathematik und Informatik) - Freien Universität Berlin, 2007.

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: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 7., 2006, Marília. *Anais...* Salvador: ANCIB, 2006.

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. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, v. 25, n. 11, p. 1251-5, nov. 2007.

SMITH, B.; KUMAR, A. On controlled vocabularies in bioinformatics: a case study in the gene ontology. *BIOSILICO: Drug Discovery Today*, v. 2, p. 246-252, 2004.

SMITH, B.; ROSSE, C. The role of foundational relations in the alignment of biomedical ontologies. In: WORLD CONGRESS ON MEDICAL INFORMATICS (MEDINFO), 11., 2004. *Proceedings...* San Francisco, CA, USA, September 7-11, 2004. In: *Studies in Health Technology and Informatics*. Amsterdam, v. 1, n. 107, p. 444-448, 2004.

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.

SWARTOUT, W. ; TATE, A. Guest editors' introduction: ontologies. *IEEE Intelligent Systems*, v. 14, n.1, p. 18-19, Jan. 1999.

GUARINO, N. Formal ontology and information systems. In: FOIS '98, 1., 1998, Trento, Italy. *Proceedings...* Amsterdam: IOS Press; Tokyo: Omsha, 1998a . p. 3-15.

VICKERY, B. C. Ontologies. *Journal of Information Science*, London, v. 23, n. 4, p. 227-286, 1997.

WEINSTEIN, P. C. Ontology-based metadata: transforming the MARC Legacy. In: ACM DIGITAL LIBRARY CONFERENCE, 3., 1998. *Proceedings...* Pittsburgh, PA, USA, June 1998.

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

WUSTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, L'informatique et les sciences des choses. In : RONDEAU, G. ; FELBER, E. (Orgs.). *Textes*

choisis de terminologie. Québec : GIRSTERM, 1981. V.I : Fondements théoriques de la terminologie, p. 57-114.