

Modelização de domínios de conhecimento: uma investigação de princípios fundamentais

Maria Luiza de Almeida Campos

Doutora em Ciência da Informação. Universidade Federal Fluminense.

E-mail: mlcampos@nitnet.com.br

Resumo

As teorias e metodologias desenvolvidas, tanto no escopo temático da ciência da informação quanto no da ciência da computação, que estão voltadas para a representação do conhecimento, apresentam, de forma mais ou menos abrangente, discussões sobre princípios como contexto de conhecimento, natureza dos conceitos, relações entre conceitos e sistemas de conceito. No campo do quadro teórico, investigam-se os modelos de abstração utilizados nas duas áreas e modelos representacionais associados à modelagem de sistemas de banco de dados, especificamente o modelo orientado a objetos. Da ciência da informação, analisam-se teorias ligadas à representação de sistemas de conceitos, como a teoria da classificação facetada e a teoria do conceito. Da teoria da terminologia, utilizam-se princípios estabelecidos para a determinação de conceitos e suas relações. Garantindo a especificidade de cada área, a comparação entre as teorias se dá, tendo em vista quatro pontos relacionados ao processo de modelização, a saber: o método de raciocínio; o objeto de representação; as relações entre os objetos; as formas de representação gráfica.

Palavras-chave

Organização do conhecimento; Modelização; Representação do conhecimento.

Modelling of knowledge domains: an investigation of fundamental pricips

Abstract

Theories and methodologies developed within the subject scope of both Information Science and Computer Science and designed to knowledge representation, present, in a somewhat comprehensive way, discussions on principles such as knowledge contexts, nature of concepts, relationships between concepts and systems of concepts. Thus, in terms of theoretical framework, abstraction models used by the two fields are investigated. As to Computer Science, representational models related to the modeling of data bank systems are investigated, more specifically the object-oriented model. Concerning information science, theories related to the representation of systems of concepts are analyzed, such as the theory of faceted classification and the theory of concept. As to the theory of terminology, principles established for determining concepts and their relationships are used. While assuring the specificity of each area, theories are compared, aiming at four topics related to the modeling process: the method of reasoning; the object of representation, the relationships among objects; the forms of graphical representation.

Keywords

Knowledge organization; Modeling; Knowledge representation.

INTRODUÇÃO

Esta investigação discute a problemática representacional, comparando os mecanismos de abstração presentes nas teorias da ciência da informação, da terminologia e da computação, que permitem a representação de conhecimento em domínios diversos, em uma proposta de encontrar um núcleo comum de conceitos imprescindíveis ao ato de modelar o conhecimento. Da ciência da informação, analisam-se teorias ligadas mais especificamente à representação de sistemas de conceitos, como a teoria da classificação facetada de S. R. Ranganathan (Ranganathan, 1951, 1967) e a teoria do conceito de I. Dahlberg (Dahlberg, 1978, 1978 a, 1983), que possibilitam a representação de domínios de conhecimento. Da ciência da computação, investigam-se os modelos representacionais associados à modelagem de sistemas de banco de dados, mais especificamente o modelo orientado a objetos (Gray, 1992; Furlan, 1998; Rumbaugh, 1994), além de um campo mais recente que repensa as possibilidades representacionais e de organização de domínios de conhecimento – a ontologia formal (Guarino, 1998, 1998a, 1994, 1997; Gruber, 1993; Sowa, 2000). Da teoria da terminologia, utilizam-se princípios estabelecidos para a determinação de conceitos e suas relações apresentadas por E. Wuester (Wuester, 1981) e seus seguidores (Drozd, 1981; Felber, 1981, 1984; Kandelaki, 1981; Riggs, 1979). Os princípios resultantes deste estudo pretendem auxiliar o modelizador em sua atividade de elaboração de modelos de representação.

Nos últimos 10 anos, desenvolvemos uma postura de pesquisa que vem privilegiando o estudo sistemático das teorias subjacentes a essas áreas, relativas à construção de modelos conceituais, que permitem a elaboração de linguagens documentárias, sistemas computacionais, hipertextos, sistemas voltados para a construção de bases de conhecimento – os chamados sistemas especialistas – e, mais recentemente, no âmbito da inteligência artificial, as ontologias (Campos, 2001; 2001 a).

Desta forma, pretendemos agregar diversas possibilidades representacionais para se pensar sobre uma dada realidade, a partir de uma postura teórico-metodológica que dê condições ao modelizador de ultrapassar modelos específicos de representação e pensar nos princípios

subjacentes ao processo de modelização. Entende-se modelização a partir dos estudos desenvolvidos por Le Moigne, em sua teoria do sistema geral, também denominada teoria da modelização (Le Moigne, 1977).

Le Moigne afirma que conhecer é modelizar, ou seja, o processo de conhecer equivale à construção de modelos do mundo/domínio a ser construído que permitem descrever e fornecer explicações sobre os fenômenos que observamos. Ele propõe o desenvolvimento de uma postura metodológica na qual possamos pensar não somente na diversidade de modelos, mas principalmente nos princípios que possibilitam o ato de modelar. Estes princípios são interessantes porque restauram a liberdade do modelizador, não o deixando refém de um dado modelo específico.

Em sua obra, Le Moigne apresenta cinco perspectivas para pensar a modelização. A primeira é instrumental: a teoria do sistema geral deve servir como uma teoria da modelização, que se fundamenta, *a priori*, postulando a pluralidade dos métodos de modelização dos fenômenos. A segunda perspectiva vai ser apresentada a partir da obra de Edgar Morin (Morin, 1990), que nos introduz ao paradigma do pensamento complexo. Em sua exposição, arquitetura um hiperparadigma da complexidade, que consiste no entendimento da modelização de todo o fenômeno percebido e concebido como complexo, pela recusa da sua simplificação, da sua mutilação, Edgar Morin (Morin, 1990) propõe que atentemos para duas “ilusões” que desviam o espírito do problema do pensamento complexo:

“A primeira é crer que a complexidade conduz à eliminação da simplicidade. A complexidade aparece certamente onde o pensamento simplificador falha, mas integra nela tudo o que põe ordem, clareza, distinção, precisão no conhecimento...; a segunda ilusão é confundir complexidade e completude, a ambição do pensamento complexo é dar conta das articulações entre domínios disciplinares, que são quebrados pelo pensamento disjuntivo (que é um dos aspectos principais deste pensamento simplificador); este isola o que aquele separa e oculta tudo o que o liga, interage, interfere. Neste sentido, o pensamento complexo aspira ao conhecimento multidimensional” (Morin, 1990, p. 8).

Na verdade, Edgar Morin quer nos alertar que a patologia moderna do espírito está na hipersimplificação que a torna cega perante a complexidade do real, enquanto o pensamento simplificador desintegra a complexidade do real. A terceira perspectiva, igualmente sugerida por Edgar Morin, diz respeito à representação e à nossa compreensão

de uma renovação da compreensão de organização, que seria a capacidade de um sistema para, ao mesmo tempo, produzir e produzir-se, ligar e ligar-se, manter e manter-se, transformar e transformar-se, concebida de um modo a não se reduzir ao conceito de estrutura. A quarta perspectiva diz respeito à distinção entre sistema e conjunto. Um sistema não é um conjunto: o conceito de sistema está fundado em uma dialética do organizado e do organizante, e não em uma enumeração de elementos e de suas relações, mas antes de tudo em uma perspectiva em que esses elementos estão dispostos uns em relação aos outros, formando o que definimos por sistema. A quinta perspectiva considera o progressivo reconhecimento da liberdade criadora do modelizador, afirmando que nenhuma ciência nos força, em nome da boa utilização da razão humana, a utilizar um, e apenas um, algoritmo de modelização. Reforça, com esta última perspectiva, que a condição fundamental para garantir a liberdade do artesão é a explicitação dos axiomas sobre os quais vai apoiar, progressivamente, suas inferências e imprimir seu projeto.

Assim, na perspectiva de propiciar ao modelizador uma gama de ferramentas representacionais para a modelização de domínios e a partir de uma investigação sistemática dos métodos e teorias, chegamos à determinação de quatro princípios fundamentais que podem ser utilizados no ato de modelar domínios de conhecimento.

O primeiro princípio diz respeito ao método de raciocínio utilizado para a organização do conhecimento dentro de um domínio. O segundo analisa como está definido o objeto de representação, ou seja, qual é a unidade de conhecimento que se vai representar. O terceiro diz respeito à relação entre os objetos, objetivando verificar as possibilidades de ligação/separação semânticas entre os conceitos de um dado domínio. O quarto evidencia as formas de representação gráfica que um modelo pode adotar.

Neste artigo, vamos apresentar uma síntese das análises desenvolvidas a partir destes quatro pontos. Antes, é necessário tratarmos de um conceito que, no nosso caso, está intrinsecamente ligado ao de modelização: o conceito de representação de conhecimento.

A REPRESENTAÇÃO DE CONHECIMENTO

O processo de modelização requer o deslocamento do “mundo fenomenal” para um espaço de representação. Segundo Davis *et alii* (1992), 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, em vez 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 à 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 selecionarmos uma representação, estamos tomando um conjunto de decisões sobre como e o que ver no mundo. Portanto, 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? (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, a representação pode, então, não ser de muita valia para o problema em mãos.

5. Uma representação de conhecimento é um meio de expressão, isto é, uma linguagem na qual se pode 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, muitas vezes não podemos, como usuários, saber se a representação não é capaz de expressar alguma coisa que gostaríamos de dizer ou se 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.

Os mecanismos de representação de conhecimento permitem, assim, que processos de formalização sobre os objetos e suas relações, em contextos predefinidos, possam ser facilmente representados. No âmbito da ciência da computação, eles servem para auxiliar a implementação de estruturas computáveis. No âmbito da ciência da informação, possibilitam a elaboração de linguagens documentárias verbais e notacionais, visando à recuperação de informação e à organização dos conteúdos informacionais de documentos. No âmbito da terminologia, esses mesmos mecanismos permitem a sistematização dos conceitos e, conseqüentemente, a elaboração de definições consistentes.

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

O nível lógico é o nível da formalização. Não existe, entretanto, preocupação com a semântica em termos dos conceitos e de suas relações; na verdade, 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; 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 deste esqueleto.

Na perspectiva da ciência da informação, da ciência da computação e da terminologia, interessam-nos aqueles níveis em que 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 estruturado e formalizado.

No nível epistemológico, como vimos, especificam-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 nível de estruturação, o nível ontológico é o nível de significação.

Na ciência da computação, modelos de objetos e de dados são exemplos de representação no nível epistemológico. Contudo, para representar conhecimento, tais representações são ainda pobres. Esses formalismos estabelecem apenas significados particulares de estruturação. É necessário, portanto, introduzir a noção de nível ontológico (Guarino, 1994). A ontologia formal, como o próprio nome indica, é um formalismo classificado no nível ontológico, pois sistematiza conhecimento pretendendo a formalização de definições axiomáticas.

Na ciência da informação, a teoria da classificação pode ser considerada em um nível de transição entre o nível epistemológico e ontológico: apesar de não pretender chegar à definição dos conceitos de um dado domínio, ela possui um formalismo que possibilita a representação do conhecimento. Já as teorias do conceito e da terminologia podem ser classificadas como de um nível ontológico propriamente, pois permitem a sistematização de conhecimentos e possuem diretrizes para a elaboração de definições.

PRINCÍPIOS PARA A MODELIZAÇÃO DE DOMÍNIOS DE CONHECIMENTO

A partir dos estudos desenvolvidos, como apresentamos anteriormente, os quatro princípios fundamentais que devem auxiliar o modelizador na elaboração de estruturas conceituais em domínios de conhecimento são o método

de raciocínio, o objeto de representação, as relações entre os objetos e as formas de representação. Nesta seção, apresentaremos cada um deles, estabelecendo um diálogo entre os domínios da ciência da informação, da ciência da computação e da terminologia.

O método de raciocínio

Modelos são elaborados, tradicionalmente, tendo como princípio um dos dois métodos de raciocínio: o método dedutivo, também denominado *top-down*, ou o método indutivo, também denominado *bottom-up*, ambos na ciência da computação.

O método indutivo possibilita a elaboração de modelos, partindo, desde o início, da representação dos elementos/objetos e relações de um contexto. Já o método dedutivo propõe que se elaborem mecanismos de abstração para pensar primeiramente o domínio/contexto, independentemente de pensar os elementos e suas relações; esta seria uma etapa posterior.

Na ciência da informação, a teoria da classificação facetada (Ranganathan, 1967) é representante de um modelo que utiliza o método dedutivo para classificar o conhecimento dentro de um domínio. Desta forma, possui mecanismos de representação para trabalhar com metaníveis conceituais – as categorias. É a partir delas que os conceitos são ordenados para formar classes de conceitos.

É interessante ressaltar que, já na década de 20 do século passado, Ranganathan utiliza o método dedutivo para a organização de uma espécie de classificação bibliográfica que vai denominar uma policotomia ilimitada. Esta espécie de estrutura de classificação se diferencia da tradicional classificação utilizada na época – a dicotômica/binária.

Os métodos de divisão, ou seja, aqueles que auxiliam a organização do conhecimento em um dado contexto, foram durante muitos séculos dicotômicos. Na dicotomia encontram-se duas divisões no primeiro estágio, duas divisões de cada uma destas divisões são formadas no segundo estágio e assim por diante. A representação esquemática da dicotomia chama-se “Árvore de Porfírio”. Ranganathan propõe uma mudança desta representação, fazendo uma analogia com um tipo de figueira indiana chamada Árvore Baniana, que se aproxima muito mais de uma árvore de classificação: do tronco original formam-se muitos outros troncos secundários, de tempos em tempos. Cada tronco foi considerado por Ranganathan como categorias. Assim, os domínios teriam um corte policotômico, e não dicotômico.

No domínio da ciência da computação, a ontologia formal utiliza o método indutivo. Apesar de possuir princípios para descrição de metaníveis de objetos em um domínio (universais), não utiliza esta classificação como um mecanismo inicial para a organização dos objetos em um contexto. O processo é iniciado com a descrição bastante específica dos objetos de um contexto, desde sua identidade até a sua dependência com outros objetos. Esta dependência, entretanto, não é estabelecida do contexto para o objeto, mas entre os objetos. Na verdade, o objetivo que suporta esta teoria é a elaboração de instrumentos que possibilitem descrever, conceitualmente, os elementos de um contexto. Desta forma, ela está mais preocupada com os significados dos conceitos do que com sua estrutura. Entretanto, acreditamos que em toda definição está implícita uma classificação. Definir é evidenciar não somente quais são os elementos constitutivos do objeto e sua funcionalidade em dado contexto, mas também o que é o objeto. Desta forma está implícita uma estrutura classificatória.

Ainda no domínio da ciência da computação, a orientação a objetos (OO), atendendo aos requisitos das aplicações a que se destina, pode ser classificada, da mesma forma, como de caráter indutivo. Nela, a metodologia já parte estabelecendo princípios para identificar os objetos de um domínio e seus relacionamentos, tratando assim, especificamente, desta representação. Esta concepção de classificação de um domínio não cabe na OO, pois seus princípios foram estabelecidos para dar conta da modelagem de dados, e não de unidades de conhecimento.

Por outro lado, é interessante observar que, na modelagem conceitual de um dado domínio, a ontologia formal avança em uma perspectiva teórica do modelo OO, pois distingue a representação do domínio, dos elementos que constituem esse domínio e de suas relações. Apresenta, por exemplo, primitivas para pensar sobre o domínio – a questão da ontologia dos universais. Em OO, verificamos que só é possível representar objetos e relações, o que não configura uma taxionomia, pois não existem elementos que possibilitem pensar a estrutura do domínio. Inicia-se o processo de modelização com os elementos e suas ligações.

Fora do domínio da ciência da computação, a teoria da terminologia pode também ser vista como aquela que suporta o método indutivo para pensar um dado contexto. Para Wuester, os conceitos (objeto/entidade/instância) associam-se um em relação aos outros, formando um sistema de conceitos. Entretanto, o sistema é formado a partir da análise do próprio conceito, e não

do contexto em que ele está inserido. O contexto é visto como um *a priori* que só é possível identificar a partir do próprio conteúdo conceitual. Na terminologia, não se fala sequer em categorias ou universais.

No âmbito da ciência da informação, por sua vez, a teoria do conceito introduz uma metodologia que poderíamos denominar 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. Em uma proposta sistêmica, que leva em conta tanto os princípios de modelização de Le Moigne (1977), quanto a 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, segundo a qual o indutivo e o dedutivo, a análise e a síntese, são vistos como processos não disjuntos, que se complementam em um todo, ou seja, “análisesínteseanálisesíntese...”

O objeto de representação

O segundo princípio diz respeito ao objeto de representação, que, nas teorias analisadas, tem sido definido, de forma geral, como a menor unidade de manipulação/representação de um dado contexto.

Na análise das diversas teorias e metodologias, identificamos, em cada uma delas, o que é considerado um objeto de representação, sua importância como a menor unidade de representação dentro de um dado contexto e as implicações que resultam desta definição.

No âmbito da ciência da informação, a teoria da classificação e a teoria do conceito se apresentam como teorias que suportam um dos primeiros esforços teórico-metodológicos que evidenciam a questão da unidade de representação. Ranganathan, em seu Prolegomena, introduz o conceito de isolado como a unidade mínima e manipulativa de um sistema de classificação. Dalhberg 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.

No âmbito da ciência da computação, no contexto da orientação a objetos, os objetos que constituem o mundo real são a sua própria representação: um objeto é um conceito. No espaço teórico da inteligência artificial, no âmbito da ontologia formal, Guarino (1998) propõe que os objetos, ou particulares, sejam classificados como concretos e abstratos. Comparativamente, na teoria do conceito, estes são considerados objetos gerais, pois ambos representam uma classe de objetos, e não uma determinada entidade no mundo, como “o cavalo branco de Napoleão”. Este, especificamente, é considerado por Dalhberg como um objeto individual. Os objetos concretos, no âmbito da ontologia formal, são classificados como contínuos e ocorrentes. Para a teoria do conceito, assim como para a teoria da terminologia, os objetos ocorrentes não são objetos, e sim ação. Uma ação não é uma entidade (objeto), mas um processo que ocorre com um dado objeto, não considerando, então, como um objeto, o que a ontologia formal denomina objeto ocorrente. Entretanto, ainda no âmbito da ontologia, Sowa (Sowa, 2000) define objeto ocorrente como um processo*. A literatura apresenta o conceito de objeto de forma conflitante.

Em geral, o que se verifica é que as teorias ligadas à ciência da computação ainda não diferenciam, de forma explícita, o mundo real (objetos) da representação deste mundo. Acreditamos que isso se explique por trabalhar com a representação de fatos, e não de unidades de conhecimento, pois o objetivo é a descrição de fatos, e não exatamente o entendimento dos conteúdos conceituais do objeto representado. Entretanto, os sistemas baseados em conhecimento, ao contrário do convencional processamento de dados, têm elementos que são mais do que dados isolados, conceitos que descrevem objetos e suas propriedades.

Estas discussões objetivam evidenciar a multiplicidade de interpretações e acabam por possibilitar uma dada flexibilidade para a observação e trabalho em contexto de domínios diferenciados. Além disso, fazem com que o modelizador sinta a necessidade de construir os pressupostos de onde irá partir ou, ainda, de tentar perceber de onde se parte para a elaboração de um modelo de representação.

* “...objeto é na atualidade considerado como contínuo,... processo é na atualidade considerado como ocorrente” (Sowa, 2000).

As relações entre os objetos: a elaboração de estruturas conceituais

As relações entre os objetos de um dado contexto formam a estrutura conceitual deste contexto e são de natureza diversa. Nesta perspectiva, diz Edgar Morin (1990), “todo modelo reflete a complexidade do real”. A análise será apresentada a partir do que vamos denominar movimentos do ato de modelar. Estes movimentos refletem grupos de relações entre conceitos.

Relação categorial

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 relação categorial aparece na teoria do conceito como relação formal-categorial; na teoria da classificação facetada, como categoria; na ontologia formal, como teoria dos universais. Nestas três teorias, a noção de categoria se coloca como um elemento agregador que reúne os conceitos em um nível mais alto de uma dada taxionomia.

Na teoria da classificação facetada, entretanto, as categorias são predefinidas, ou seja, Ranganathan propõe que todo contexto de conhecimento pode ser dividido em até cinco categorias fundamentais: entidade, matéria, energia, espaço e tempo. O método parte, então, da observação do universo de conhecimento/discurso em que se está inserido, de sua análise, determinando dentre as cinco categorias aquelas que representam aquele contexto, para só depois inserir os conceitos que fazem parte de cada categoria*.

A teoria do conceito estende para um número muito maior as categorias. Na verdade, Dalhberg explicita, em seu trabalho *Ontical Structures* (Dahlberg, 1978a), que tem por princípio a classificação proposta por Aristóteles na antigüidade clássica, categorias para o entendimento do conceito, e não para classificar um domínio de conhecimento/atividade (ver quadro 1, a seguir). Entretanto, como é possível perceber, o entendimento do conceito está refém de um contexto que é classificado

* O Classification Research Group (CRG), criado na Inglaterra para desenvolver estudos com base nos princípios de Ranganathan, identificou grande diversidade de manifestação das cinco categorias (Wilson, 1972).

por conta dos conceitos que nele estão inseridos. Desta forma, o contexto e seus elementos estão em uma relação unívoca. Além deste ponto diferente, mas não divergente de finalidade, Dalhberg apresenta o conceito de propriedade como uma categoria, diferentemente de Ranganathan, que o coloca como uma manifestação de uma das categorias. Por exemplo: para Ranganathan, toda propriedade está refém de uma das categorias. Assim, no contexto da engenharia civil, o concreto é um material de construção que está na categoria material que possui a propriedade de cisalhar, ou seja, o cisalhamento não ocorre como uma manifestação isolada, sendo refém de uma entidade ou processo no mundo fenomenal.

Na ontologia formal, o conceito categoria aparece como o nível de análise da estruturação dos objetos de um domínio. Na verdade, Guarino parte dos objetos (ontologia dos particulares) e nos apresenta a seguinte forma de análise: em todo contexto nos deparamos com objetos (contínuos ou ocorrentes) que estão em dada região temporal ou espacial. Quando esses objetos fazem parte de um dado domínio, eles possuem propriedades e estão relacionados entre si. Os tipos de objetos, as categorias de objeto e os papéis que esses objetos exercem em um dado domínio são propriedades utilizadas para análise dos objetos em determinado contexto. A categoria, especificamente, é considerada uma classe de nível mais amplo, tendo como função possibilitar uma classificação geral do domínio em questão. Neste ponto, encontramos semelhança entre Ranganathan, Dalhberg e Guarino. Entretanto, Guarino não nos apresenta que categorias seriam estas; somente aponta para a necessidade de se classificar um domínio em categorias.

A teoria da terminologia e a orientação a objetos não possuem esta relação. No que diz respeito à teoria da terminologia, o maior nível de agrupamento de conceitos é o que ela denomina sistema de conceitos, que possui um significado próprio, pois, aqui, o sistema de conceitos representa uma classe de conceitos, e não a reunião de todas as classes, que só seria possível se existisse um elemento agregador que, no caso, podem ser as categorias. Isto acaba não acontecendo, porque a teoria da terminologia utiliza um método indutivo partindo de cadeias ascendentes para a definição do conceito, e o contexto é relevante somente para a definição do conceito, não para possibilitar uma representação do domínio. A orientação a objetos parte do objeto e de sua relação, mas não está preocupada exatamente em classificar o objeto no contexto de categorias preestabelecidas, e sim no contexto em que a aplicação se insere, ou seja, descrevê-lo e apresentá-lo na relação com outros objetos.

QUADRO 1
Correlação das categorias de Aristóteles com as categorias de Dahlberg

Aristóteles		Categorias conceituais de Dahlberg
Substância	ENTIDADES	Fenômenos Objeto geral Objeto material
Quantidade Qualidade Relação	PROPRIEDADES	Contagem e mensuração Qualidade Comparação
Tempo Posição Espaço	DIMENSÕES	Tempo Posição Espaço

Relação hierárquica

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.

Neste sentido, a teoria do conceito e a teoria da classificação denominam esta relação de hierárquica, a teoria da terminologia denomina relação lógica, já a orientação a objetos define como relação de generalização e especialização, e a ontologia formal define como “é um (ISA)” e é considerada uma propriedade em uma ontologia mínima de universais para a estruturação de um domínio.

Na teoria do conceito, a relação hierárquica é de dois tipos: a relação de abstração de gênero e espécie, que forma cadeias de conceitos, e a relação lateral, ou seja, que forma renques de conceitos. Como a teoria da classificação, a teoria do conceito utiliza a terminologia relação hierárquica. Entretanto, para Ranganathan, uma relação hierárquica possui conceitos que não estão somente em uma relação de generalização e especialização, mas também em uma relação parte-todo. Para Dalhberg, a relação hierárquica baseia-se em uma relação lógica de implicação, ou seja, nela os conceitos devem ser da mesma natureza, o que não ocorre com o todo e suas partes, que, em muitos casos, são de natureza diferentes. Por exemplo: a instalação hidráulica de um edifício faz parte dele, mas não é da mesma natureza do edifício, ou seja, não é um tipo de edifício.

Na teoria da terminologia, o que Dalhberg denomina hierárquica Wuester classifica de relação lógica e específica muito mais os seus tipos - relação de comparação (subordinação lógica, coordenação lógica, interseção lógica e diagonal) e relação de combinação (determinação, conjunção, disjunção). Para a terminologia, esta relação é fundamental para a elaboração de definições, pois permite a compreensão do conceito, isto é, especifica o que a “coisa é”. Como, entretanto, para a terminologia, o objetivo de evidenciar as relações entre os conceitos não é para especificar uma estrutura sistemática, mas para o entendimento do próprio conceito, acreditamos que a noção de hierarquia, como se apresenta, não satisfaz. Na verdade, esta noção de hierarquia aparece nas teorias de representação de unidades de conhecimento da ciência da informação, por meio de autores, como Ranganathan, que estavam preocupados em elaborar sistema de classificação para a guarda de documentos, na qual seria necessária uma certa noção de precedência, do qual acreditamos que venha a aplicação do termo. Dalhberg, ao se debruçar na área de documentação para a elaboração de tesouros, utilizou a noção de hierarquia no sentido de precedência, para representar a generalização e especialização.

Na orientação a objetos, a relação equivalente é denominada generalização e especialização; mas, aqui, só se considera a relação que, na teoria do conceito, é denominada abstração, isto é, a que forma cadeias - macieira, por exemplo, é uma árvore frutífera que, por sua vez, é uma árvore -, e não a que forma renques - macieira, parreira e laranjeira são tipos de árvore frutífera. Este tipo de relação é fundamental em OO, pois dá suporte ao mecanismo de reuso, onde novas classes são criadas a partir de classes preexistentes, especializando um conceito.

Na ontologia formal, a relação equivalente é considerada somente como de gênero e espécie e está no escopo da estruturação de um domínio de conhecimento. Esta relação é considerada, na ontologia formal, como a principal para a organização de uma taxionomia. Um aspecto importante a ressaltar na ontologia formal é o conceito de teoria da identidade, no qual são introduzidas questões que nos levam a perceber como um conceito pode ser considerado um ente no mundo, o que permite definir, através de um “pensar” classificatório, aquilo que pode ser considerado gênero e espécie de conceito, fato que não se verifica nas outras teorias e métodos.

A relação hierárquica é uma das relações principais em qualquer estrutura classificatória. Ela é a que forma a espinha dorsal de uma estrutura. Para as teorias que

tratam dos processos definitórios de um conceito, como a teoria da terminologia e a ontologia formal, ela é imprescindível, sendo a partir dela que se estabelece o primeiro elemento de uma definição. Por exemplo, na definição do conceito “homem”, iniciamos com outro conceito superordenado “animal racional”.

Relação partitiva

Outro 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.

Como as outras relações, esta é denominada diferentemente nas teorias e métodos analisados. A teoria do conceito a denomina relação partitiva; a teoria da classificação coloca, em um mesmo grupo, as relações hierárquicas e partitivas, como relação hierárquica; na teoria da terminologia, ela é chamada de relação ontológica de coordenação partitiva; a orientação a objetos a define como relação de agregação; e a ontologia formal, de teoria todo-parte, também denominada “mereotopologia” (*mereotopology*).

Com exceção da ontologia formal e da terminologia, as outras teorias e métodos tratam esta relação de forma muito simples, como a parte de um todo. Entretanto, não especificam o que é o todo ou o que é a parte. Na verdade, é somente a ontologia formal que apresenta critérios que esclarecem esta questão, pois a terminologia apresenta a tipologia das relações partitivas a partir das possibilidades de coordenação e interseção entre os objetos. Esses critérios auxiliam o modelizador a distinguir a natureza daquilo que irá ser considerado como parte e como todo; funcionam também como “metacategorias informacionais” para identificar as relações partitivas de um dado domínio. Como exemplo do que pode ser considerado “metacategorias informacionais” partitivas, apresentamos:

- objeto integral/componente - o que caracteriza este tipo de partição é a possibilidade de a estrutura do componente ser separável e possuir uma funcionalidade específica, no caso da roda do carro;
- membro/coleção - o componente não possui nenhum papel funcional em relação ao todo, isto é, são partes do todo, mas podem ser individualizados, no caso da árvore como parte da floresta;
- massa/porção - o todo é considerado como um agregado homogêneo, e sua parte/porção é similar e separável, no caso do pedaço de uma torta;

- objeto/matéria – expressa a parte/ matéria na qual a coisa é feita e que não é separável do objeto, pois não tem nenhuma função separada do objeto, no caso da bicicleta e do material de que ela é feita – o aço;
- caráter/atividade – designa uma parte de uma atividade, no caso da indexação que faz parte da fase de tratamento documentário;
- área/lugar – é um tipo partitivo de relação espacial entre uma região ocupada por diferentes objetos, no caso o oásis e o deserto.

Relações entre categorias

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 (lógicas e partitivas), podem ser reconhecidas como relações que tornam evidente uma determinada demanda, ou função, entre os objetos no mundo fenomenal, não objetivando explicitar o objeto e suas propriedades.

Neste sentido, Dahlberg apresenta, em sua teoria, algumas das seqüências possíveis para se estabelecer uma relação funcional sintagmática entre os conceitos. Entretanto, diferentemente da teoria da terminologia, Dahlberg não esgota os tipos de relacionamento entre categorias, porque coloca esta relação como aquela que deve solucionar demandas no mundo fenomenal e, como este “mundo” é dinâmico, considera impossível esgotar as possibilidades de relacionamento. Se, por um lado, esta opção é mais flexível, por outro, provoca algumas inconsistências, pois não se define claramente que tipo de objeto pode estar relacionado a outro, no momento de se constituir um relacionamento.

Já a teoria da terminologia procura disciplinar a possibilidade representacional deste tipo de relação, ou seja, propõe uma classificação para o que ela irá denominar relação ontológica de encadeamento e relação ontológica de causalidade, diferenciando, assim, a relação ontológica de coordenação partitiva. Na relação de encadeamento, estabelece-se a possibilidade da verificação de existência, em um dado contexto, de um relacionamento de contigüidade no tempo, como, por

exemplo, predecessor-sucessor. Na relação de causalidade, verifica-se a possibilidade de estabelecimento de um elo sucessivo de causas: a relação genealógica (pai-filho); a relação ontogenética (ovo-larva); a relação de substância (urânio II-rádio). Além desses, Wuester descreve outros relacionamentos entre categorias como material – produto e instrumental (instrumento e seu uso).

A ontologia formal não possui uma classificação específica para o relacionamento entre categorias. Entretanto, ao definir os pressupostos da teoria da dependência, apresenta questões ligadas ao relacionamento entre os conceitos, no qual evidencia, com exemplificações, que os conceitos se relacionam não somente em dependência genérica e partitiva, mas também em várias formas de dependência existencial, envolvendo indivíduos específicos pertencentes a classes diferentes. Guarino cita, inclusive, um exemplo do que ele irá denominar dependência de classe – o relacionamento entre pai e filho – que Wuester, na teoria da terminologia, classifica como um tipo de relacionamento entre categorias, denominado relação genealógica.

A teoria da classificação facetada não apresenta esta relação, pois ela somente vai trabalhar com relações que evidenciam estruturas hierárquicas (formam cadeias e renques), o que não é o caso das relações entre categorias. Desta forma, este tipo de relação não se coloca em estruturas de classificação para guarda de documentos.

A orientação a objetos não detalha esta relação, mas se verifica que ela existe como um meio de associação entre os objetos. Na verdade, a orientação a objetos não nomeia esta relação, mas a utiliza quando efetiva relacionamentos entre objetos. Quando se diz, por exemplo, em uma dada implementação, que aluno cursa disciplina, a relação cursa não pode ser considerada uma generalização/especialização nem uma agregação; é considerada, então, uma associação entre os objetos. Apesar de o conceito de associação na orientação a objetos englobar também generalização/especialização e agregação, ele contém outros relacionamentos que podem ser considerados entre categorias.

Relação de equivalência

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, dá-se no âmbito da língua: a chamada relação de equivalência. Esta relação é somente representada na teoria do conceito e da terminologia. Entretanto, na teoria da classificação

facetada, Ranganathan (1967) cita a questão da linguagem em um plano que ele denomina verbal, ou seja, está no âmbito da linguagem/comunicação, e não da representação conceitual, estando em um plano que ele denomina das idéias e, por isso, não se coloca. Em uma visão comparativa, acreditamos que a questão das relações de equivalência não se coloca na orientação a objetos por este mesmo motivo, ou seja, o que se deseja é a representação de um contexto de atividade, e não estabelecer meios de comunicação entre “grupos de falantes”, como é o objetivo da teoria da terminologia. Entretanto, esta é uma conclusão que consideramos apropriada para justificar a inexistência deste relacionamento, pois a literatura de OO nem toca neste ponto.

No que concerne à base teórica da ontologia formal, esta relação não se apresenta. Entretanto, as ontologias, quando construídas, estabelecem sinonímias entre as diversas formas de denominação. Acreditamos que, como esta relação não faz parte de um plano lógico, mas de um plano da língua, ela não tem sido apresentada nos modelos que compõem as relações entre os objetos.

As formas de representação gráfica

Um modelo conceitual deve ser visto, também, como um espaço comunicacional em que transpomos o mundo fenomenal para um espaço de representação. Desta maneira, consideramos importante investigar formas gráficas de representação, porque, em alguns casos, estas formas “inibem” as possibilidades representacionais.

Nas teorias e metodologias que nos propomos a analisar, verificamos que 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 quanto na teoria do conceito, não é apresentado nenhum modelo para expressar graficamente as relações conceituais. Acreditamos que isso se deva ao fato de que essas teorias têm por objetivo a elaboração de linguagens documentárias que, apesar de possuírem uma parte sistemática, não têm os conceitos representados em forma gráfica, mas em forma de uma lista endentada de termos, com uma notação que, de certa forma, deixa evidentes os grupos de termos afins.

Na área da construção de terminologias, em que se configuram a teoria da terminologia e a ontologia, as representações gráficas são apresentadas de forma consistente, entretanto com algumas diferenças fundamentais. A teoria da terminologia utiliza-se de uma representação gráfica, visando a dois aspectos: o primeiro fornece ao terminólogo uma ferramenta mais precisa para classificar os conceitos em um sistema de conceitos, possibilitando, assim, a elaboração de definições mais consistentes; o segundo permite ao usuário visualizar as classes de conceitos afins, o que uma simples ordem alfabética não permitiria. Por outro lado, a ontologia utiliza-se da representação gráfica, como uma ferramenta para garantir um projeto lógico mais bem estruturado de um sistema e, para tanto, recorre geralmente à representação gráfica utilizada pelo modelo OO. Assim, a ontologia, apesar de avançar quanto aos aspectos teóricos da OO, ainda está refém de uma forma de evidenciar os conceitos e suas relações, pois, como foi dito, utiliza o modelo de representação da OO.

Destacamos que as representações gráficas carecem, muito mais que qualquer outro ponto analisado, de um espaço maior de investigação.

CONSIDERAÇÕES FINAIS

Este é um estudo em desenvolvimento. Estamos à procura da determinação de uma metodologia mais adequada para a representação de domínios de conhecimento que agregue objetivos de aplicação ligados tanto à área da ciência da informação, quanto à da ciência da computação e à da terminologia.

A proposta dos estudos desenvolvidos vem nos colocar diante das seguintes questões: a necessidade de investigar os princípios que regem as formas de pensamento; uma ação de pensar a complexidade do real; a possibilidade de liberdade na ação de modelizar; a multiplicidade das ferramentas representacionais; a garantia de criação e escolha para o modelizador.

O estudo vem, assim, apontando para o estabelecimento de maior diálogo entre as áreas da ciência da informação, da ciência da computação e da terminologia, visando a somar esforços teórico-metodológicos no sentido de fortalecer o processo de modelização.

Artigo recebido em 18-02-2004 e aceito para publicação em 24-04-2004.

REFERÊNCIAS

- CAMPOS, Maria Luiza de Almeida. *A organização de unidades de conhecimento: o modelo conceitual como espaço comunicacional para a realização da autoria*. 2001. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.
- _____. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói, RJ : Eduff, 2001a.
- DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. *International Classification*, v. 5, n. 3, p. 142-150, 1978.
- _____. *Ontical structures and universal classification*. Bangalore : Sarada Ranganthan Endowment, 1978a. 64 p.
- _____. *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, Milan, Italy. Proceedings... Milan, Italy, 1992.
- DROZD, L. Some remarks on a linguistic theory. In: *THEORETICAL AND METHODOLOGICAL PROBLEMS OF TERMINOLOGY*, 1979, Moscow. Proceedings... Muenchen : Saur, 1981.
- FELBER, H. The Vienna School of Terminology: fundamentals and its theory. In: *INTERNATIONAL SYMPOSIUM ON THEORETICAL AND METHODOLOGICAL PROBLEMS OF TERMINOLOGY*, 1979, Moscow. Proceedings... Muenchen : Saur, 1981.
- _____. *Terminology manual*. Paris : UNESCO, 1984. 234 p.
- FURLAN, José Davi. *Modelagem de objetos através da UML: the unified modeling language*. São Paulo : Makron Books, 1998.
- GRAY, Peter M. D. et al. *Object-oriented databases: a semantic data model approach*. Henselhemstead : Prentice Hall International United Kingdom, 1992.
- GRUBER, T. R. *Toward principles for the design of ontologies used for knowledge sharing*. [S. l.] : Stanford University, Knowledge Systems Laboratory, 1993.
- GUARINO, Nicola. Formal ontology and information systems. In: *FORMAL ONTOLOGY AND INFORMATION*, 1998, Trento, Italy. Proceedings... Trento, Italy : [s. n.], 1998a. p. 3-15.
- _____. The ontological level. In: *Philosophy and the cognitive science*. Vienna : Holder-Pichler-Tempsky, 1994. p. 443-456.
- _____. Semantic matching: formal ontological distinction for information organization, extraction, and integration. In: PAZIENZA, M. T. (Ed.). *Information extraction: a multidisciplinary approach to an emerging information technology*. [S. l. : s. n.], 1997.
- _____. Some ontological principles for designing upper level lexical resources. In: *INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVOLUTION*, 1. 1998. Proceedings... Granada, Spain, [s. n.], 1998.
- KANDELAKI, T. L. Les sens des termes et les systèmes desens des terminologies scientifiques et techniques. In: RONDEAU, G.; FELBER, H. *Textes choisis de terminologie: fondements théoriques de la terminologie*. Québec : GIRSTERM, 1981.
- LE MOIGNE, Jean-Louis. *A teoria do sistema geral: teoria da modelização*. Lisboa : Instituto Piaget, 1977.
- MORIN, Edgar. *Introdução ao pensamento complexo*. 2. ed. Lisboa : Instituto Piaget, 1990.
- MORIN, Edgar, LE MOIGNE, Jean-Louis. *A inteligência da complexidade*. São Paulo : Petrópolis, 2000.
- RANGANATHAN, S. R. *Prolegomena to library classification*. Bombay : Asia Publishing House, 1967. 640 p.
- _____. *Philosophy of library classification*. New Delhi : Einar Munksgaard, 1951.
- RIGGS, F. W. A. New paradigm for social science terminology. *International Classification*. v. 6, n. 3, p. 150-157, 1979.
- RUMBAUGH, J. et al. *Modelagem e projetos baseados em objetos*. Rio de Janeiro : Campus, 1994.
- SOWA, John F. *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove : Brooks/Cole, 2000.
- WUESTER, 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, F. (Org.). *Textes choisis de terminologie*. I. Fondements théoriques de la terminologie. Québec : GIRSTERM, 1981. p. 57-114.