

Roteiro para construção de uma ontologia bibliográfica através de ferramenta automatizada

Maurício B. Almeida

Mestre em Ciência da Informação pela ECI/UFMG
Professor assistente da PUC Minas

O número crescente de fontes de dados disponíveis torna cada vez mais difícil a seleção, aquisição e combinação de dados. Dentre as propostas da literatura, um grande número utiliza ontologias como forma de organizar e caracterizar dados. O uso de lógicas terminológicas baseadas em formalismos de representação do conhecimento permite a construção e implementação das ontologias em computadores. Inclui uma breve pesquisa sobre metodologias, ferramentas e linguagens para a construção de ontologias e um roteiro para a construção de ontologias através de uma ferramenta automatizada, que utiliza uma linguagem baseada em lógica OIL (*Ontology Interchange Language*). Apresenta exemplo de uso da ferramenta OIEd para construção de uma ontologia sobre referências bibliográficas. Estima que o trabalho seja um exemplo, para fins didáticos, sobre a construção de ontologias.

Palavras-chave: Ontologias; Linguagens; Metodologias; Ferramentas

Recebido em: 03.06.2003

Aceito em: 22.10.2003

1 Introdução

Para a ciência da informação as pessoas têm a responsabilidade de processar o conhecimento e, assim, uma de suas preocupações é a padronização da terminologia utilizada para se encontrar e se classificar a informação. Daí deriva a importância do uso de ontologias para caracterizar e relacionar entidades em um domínio do conhecimento.

Definir ontologias é classificar em categorias aquilo que existe em um mesmo domínio do conhecimento. Uma ontologia é um *catálogo de tipos de coisas*, as quais assume-se existir em um domínio de interesse, na perspectiva de uma pessoa que usa uma linguagem (Sowa, 1999). Historicamente, o termo ontologia tem origem no grego *ontos*, ser e *logos*, palavra. O termo original é a palavra aristotélica *categoria*, usada para classificar e caracterizar alguma coisa. Os tratados de Aristóteles estudam categorias que podem classificar entidades e introduzem o termo *differentia para* propriedades que distinguem diferentes espécies do mesmo gênero. A técnica de herança é o processo de mesclar *differentias* definindo categorias por gênero e diferenciação.

Outra contribuição aristotélica foi a invenção dos *silogismos* como um padrão formal para representar *regras de inferência*. Esses estudos são a base para as regras de herança e a fundação da *lógica de primeira ordem* que é a base da *lógica descritiva*, da qual uma das versões modernas é a OIL (*Ontology Interchange Language*) (Sowa, 1999). A OIL é uma linguagem para modelagem de primitivas que se baseia em conceitos, classes e atributos. Permite a construção e representação de ontologias, tem semântica formal bem estabelecida e tem similaridades com as linguagens para a Web (Fensel et al., 2001). O roteiro para a construção de ontologias apresentado neste trabalho, baseia-se no uso da ferramenta OIEd desenvolvida na *Manchester University*, a qual utiliza a linguagem OIL.

A seção 2 apresenta as ontologias, suas características comuns, sua tipologia e seu papel na organização da informação. A seção 3 apresenta uma pesquisa sobre a construção de ontologias, listando metodologias, ferramentas e linguagens. Na seção 4, desenvolve-se um roteiro para a construção de ontologias, através de exemplo de construção de uma ontologia bibliográfica com a ferramenta OIEd. Finalmente, a seção 5 apresenta as conclusões e aponta direções para trabalhos futuros.

2 Ontologias

Para facilitar a compreensão do que são as ontologias e sobre como podem ser utilizadas, essa seção apresenta sua definição, características comuns e seus tipos, além de considerações sobre o seu uso nas tarefas de organização da informação.

2.1 Definição de ontologia

Muitos trabalhos sobre ontologia são encontrados na literatura e são apresentadas diversas definições para o termo como as de Guarino e Giaretta (1995), Guarino (1996), Gruber (1996), Albertazzi (1996), Uschold e Gruninger (1996), Neches et al. (1991), Chandrasekaran, Johson, Benjamins (1999).

A definição de Borst (1997, p. 12), por ser simples e completa, será a adotada neste artigo: “*uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada*”. Nessa definição, *formal* significa legível para computadores; *explícita* diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas que são explicitamente definidos; *compartilhado* é o conhecimento consensual; e, *conceitualização* diz respeito a um modelo abstrato de algum fenômeno do mundo real.

As ontologias podem ser classificadas de acordo com o grau de formalidade de seu vocabulário, com a estrutura e o assunto da conceitualização, com sua função e com a sua aplicação. As características gerais de uma ontologia são apresentadas na seção seguinte, através de uma breve revisão de literatura.

2.2 Características gerais e tipologia para ontologias

As ontologias não têm sempre a mesma estrutura, mas algumas características e componentes estão presentes na maioria delas. Para Gruber (1996), os componentes básicos de uma ontologia são as *classes*, as quais são organizadas em uma taxonomia, as *relações*, que representam a interação entre os conceitos, os *axiomas*, que representam sentenças verdadeiras e as *instâncias*, que representam dados.

Uma ontologia sempre inclui um vocabulário de termos e a especificação de seu significado. O grau de formalidade com que o vocabulário é criado, bem como seu significado, classificam as ontologias como *altamente informais*, que são expressas em linguagem natural; *semi-informais*, que são expressas em linguagem natural, de forma restrita; *semi-formais*, que são expressas em linguagem artificial, definida formalmente, e as *rigorosamente formais*, que são definidas com semântica formal, teoremas e provas (Uschold e Gruninger, 1996). A partir da observação de grupos de ontologias, pode-se verificar a existência de tipos bem definidos. Pesquisas têm sido desenvolvidas no sentido de estabelecer uma classificação mais objetiva para as ontologias e de entender o tipo de conhecimento que pode ser inserido em suas classes.

Um estudo abrangente sobre os tipos de ontologias foi desenvolvido por Van-Heijst, Schreiber e Wielinga (2002), que organizaram uma biblioteca de ontologias reutilizáveis na área médica e classificaram as ontologias em relação à conceitualização, ou seja, estrutura e assunto. Distinguem-se as seguintes categorias: *ontologias terminológicas*, que especificam termos usados para representar o conhecimento no domínio; *ontologias de informação*, que especificam a estrutura de registros de bancos de dados; *ontologias de modelagem do conhecimento*, que especificam conceitualizações e são refinadas para uso no domínio que descrevem; *ontologias de aplicação*, que contêm definições para modelagem do conhecimento em uma aplicação; *ontologias de domínio*, que expressam conceitualizações para um domínio; *ontologias genéricas*, que expressam conceitualizações genéricas; e *ontologias de representação*, que explicam as conceitualizações dos formalismos de representação do conhecimento.

Mizoguchi, Vanwelkenhuysen e Ikeda (1995) também definem classes de ontologias: as *ontologias de domínio*, que são reutilizáveis e fornecem vocabulário sobre conceitos e relacionamentos; as *ontologias de tarefa*, que fornecem um vocabulário sistematizado, especificando tarefas que podem estar

no mesmo domínio; *as ontologias gerais/comuns*, que incluem vocabulário relacionado a coisas, eventos, tempo, espaço, comportamento, funções etc.

Para Haav e Lubi (2001) *ontologias de alto nível* descrevem conceitos gerais, que podem estar relacionados a todos os termos da ontologia e são independentes do domínio. As ontologias de domínio e as ontologias de tarefa descrevem, respectivamente, o vocabulário relacionado a um domínio (como, por exemplo, medicina ou automóveis) ou uma tarefa ou atividade (como, por exemplo, diagnósticos ou compras), através da inserção de termos especializados na ontologia. Jasper e Uschold (1999) identificam quatro categorias de ontologias em sistemas computacionais; a *ontologia de autoria neutra*, em que um aplicativo é escrito em uma linguagem e depois convertido para uso em diversos sistemas; a *ontologia como especificação*, em que se cria uma ontologia para desenvolvimento de *softwares*; a *ontologia de acesso comum a informação*, que torna a informação inteligível, proporcionando conhecimento compartilhado dos termos ou inter-relacionando grupos de termos.

Mesmo sem um consenso, observa-se que os tipos apresentados guardam semelhanças. Conhecidos os principais tipos pode-se buscar as ontologias já desenvolvidas e mais adequadas à tarefa a ser executada. A seção seguinte apresenta a importância da utilização das ontologias.

2.3 O papel das ontologias na organização do conhecimento

Os computadores, atualmente as principais ferramentas de disseminação de dados, não são mais dispositivos isolados e se transformaram em *pontos de acesso* inseridos em redes, caracterizadas por um intercâmbio constante (Fensel et al., 2001). Soluções que propiciam esse intercâmbio, ou inter-operabilidade, são essenciais no âmbito das tecnologias informacionais.

As ontologias podem atuar sobre fontes de dados, proporcionar organização e, conseqüentemente, recuperação mais efetiva. Possibilitam uma compreensão comum e compartilhada de um domínio, onde ocorre interação entre pessoas e sistemas. Desempenham um papel importante no intercâmbio de informações ao proporcionar estrutura semântica às fontes de dados. A comunicação entre os agentes envolvidos nos processos (computadores ou pessoas) se torna mais efetiva ao serem reduzidas diferenças conceituais ou terminológicas.

Além de seu uso no intercâmbio entre fontes de dados, as ontologias são utilizadas hoje em diversas áreas para organizar conhecimento como processamento de linguagens naturais (Bateman, 1996; Aguado et al., 1998); gestão do conhecimento (Motta, Buckingham-Shum, Domingue, 2000; Kalfoglou, 2001; Hasman et al., 1999; Domingue et al., 1998; Gandon, 2001); comércio eletrônico (Leger et al., 2000; Vázquez, Valera, Bellido, 2001); recuperação da informação na *Web* (Martin, Eklund, 2001; Alexaki et al., 2002; Borgo et al., 1997).

A existência de ontologias pode ser um fator determinante na organização e recuperação da informação em um domínio. A construção de ontologias tem sido estudada por diversos autores na literatura. A seção seguinte apresenta algumas iniciativas que visam facilitar a construção de ontologias.

3 Construção de ontologias

Existem situações em que é necessário construir ontologias:

“para compartilhar a estrutura de informação comum entre pessoas ou softwares, para reutilizar o conhecimento do domínio, para tornar explícitos fatos consensuais, para separar um domínio do conhecimento do conhecimento operacional, para analisar um domínio.” (Noy e Guinness, 2001, p. 1).

As técnicas para construção de ontologias têm se desenvolvido por iniciativas de especialistas de diversas áreas. Muitas comunidades científicas utilizam ontologias: a de inteligência artificial (no desenvolvimento de *sistemas baseados em conhecimento*), a que estuda linguagens naturais (para caracterizar o sentido e o significado das palavras), a de bancos de dados (como esquema conceitual e para buscar inter-operabilidade semântica), a de recuperação da informação (abordagem semântica), e a comunidade de ciência da informação (por exemplo, no estudo de metadados, visando a melhoria da eficiência na recuperação da informação).

Nas seções seguintes, apresenta-se uma breve revisão de literatura sobre as metodologias, ferramentas e linguagens para a construção de ontologias.

3.1 Metodologias disponíveis para a construção de ontologias

A construção e manipulação de ontologias tem sido sistematizada por metodologias. Baseado nos estudos de Corcho, Fernández-López e Gomez-Pérez (2001), apresentam-se várias metodologias para construção de ontologias. A FIG. 2 sintetiza os comentários sobre cada metodologia, proporcionando uma visão geral de seu funcionamento.

Metodologia	Breve descrição
Cyc	Codifica o conhecimento das fontes e, quando já existe conhecimento suficiente na ontologia, novo consenso é obtido por ferramentas de linguagem natural. (Lenat; Guha, 1990).
Unschold	Identifica o propósito, os conceitos e os relacionamentos entre os conceitos, além dos termos utilizados para codificar a ontologia e, em seguida, documentá-la. (Unschold; King, 1996).
Grüninger	Método formal que identifica cenários de uso da ontologia, utiliza linguagem natural para determinar o escopo da ontologia, extrai dados sobre os conceitos, propriedades, relações e axiomas, os quais são formalmente definidos em PROLOG (Grüninger; Fox, 1995).
KACTUS	Método recursivo, que consiste em uma proposta inicial para uma base de conhecimento; quando é necessária uma nova base em domínio similar, generaliza-se a primeira base em uma ontologia, adaptada a ambas aplicações; quanto mais aplicações, mais genérica a ontologia. (Bernaras; Laresgoiti; Corera, 1996 citado por Gomez-Perez, (1999).
<i>Methontology</i>	Constrói-se uma ontologia por reengenharia sobre outra ontologia; a construção utiliza o conhecimento do domínio e as atividades principais são: especificação, conceitualização, formalização, implementação e manutenção (Fernández-López et al., 1999).

Metodologia	Breve descrição
<i>Sensus</i>	Abordagem que deriva ontologias a partir de outras; identificam-se termos relevantes para o domínio, os quais são ligados à ontologia mais abrangente (<i>Sensus</i> , com 50.000 conceitos); os termos relevantes são selecionados e um algoritmo monta a estrutura hierárquica do domínio. (Swartout et al., 1996).
<i>On-to-knowledge</i>	Auxilia a administração de conceitos em organizações; identifica metas para as ferramentas de gestão do conhecimento e utiliza cenários e contribuições dos provedores / clientes de informação da organização. (Staab et al., 2001).
CO4 (<i>Cooperative Construction of Consensual knowledge bases</i>)	Permite que usuários discutam sobre o conhecimento introduzido em bases de conhecimento, as quais são compartilhadas; essas bases são ontologias, pois deve haver consenso sobre o conhecimento ali representado. Quando uma mudança é proposta, os usuários são notificados e podem aceitá-la ou não (Euzenat, 1996).
(KA)	Modela formas de aquisição do conhecimento usando ontologias desenvolvidas em conjunto por pessoas em diferentes locais, mas que utilizam o mesmo padrão; a comunicação e a coordenação são feitas via agentes inteligentes (Kietz; Maedche; Volz, 2000).
Maedche	Uma ontologia genérica é convertida em modelo; especificam-se textos e obtém-se conceitos do domínio a partir das fontes disponíveis; removem-se conceitos genéricos, de forma que apenas os conceitos específicos do domínio permaneçam; a estrutura conceitual da ontologia está estabelecida e obtém-se as relações.

3.2 Ferramentas disponíveis para a construção de ontologias

Por ser tratar de uma tarefa dispendiosa, qualquer apoio na construção de ontologias pode representar ganhos significativos. Exemplos de ferramentas para a construção de ontologias são apresentadas a seguir, na FIG. 2. A lista de ferramentas apresentada não tem a pretensão de ser completa e, apesar da preocupação em cobrir as mais representativas descritas recentemente na literatura, este estudo não é exaustivo.

Metodologia	Breve descrição
CODE4 (<i>Conceptually Oriented Description Environment</i>)	Ferramenta de propósito geral. Possui diferentes modos de herança e inferência, interface gráfica de fácil uso, modo de hipertexto para navegação e utilitários para leitura de documentos e gerenciamento léxico. (Skuce, 1995)
VOID	Ambiente para navegação, edição e gerenciamento de ontologias. Através de simulações, possibilita o estudo de questões teóricas como: organização de bibliotecas de ontologias e tradução entre diferentes formalismos. (Schreiber; Terpstra; Sisyphus, 1995)
<i>IKARUS (Intelligent Knowledge Acquisition and Retrieval Universal System)</i>	Explora as capacidades cooperativas do ambiente Web. Utiliza uma representação hierárquica gráfica que permite herança múltipla. As declarações que contém a informação são representadas como predicados, com sintaxe e semântica definidos, ou como fragmentos sem estrutura. (Skuce, 1996)
Ontolinguá	Conjunto de serviços que possibilitam a construção de ontologias compartilhadas entre grupos. Permite acesso a uma biblioteca de ontologias, tradutores para linguagens e um editor para criar e navegar pela ontologia. Editores remotos podem editar ontologias usando protocolos. (Farquhar; Fikes; Rice, 1996)

Metodologia	Breve descrição
<i>Ontosaurus</i>	Consiste de dois módulos: um servidor de ontologias, que usa o LOOM para representação do conhecimento, e um servidor de navegação por ontologias, que cria páginas HTML dinamicamente e apresenta a hierarquia da ontologia. (Swartout et al., 1996)
<i>GKB-Editor (Generic Knowledge Base Editor)</i>	Ferramenta para navegação e edição de ontologias através de sistemas de representação baseados em <i>frame</i> 's. Oferece interface gráfica, onde os usuários podem editar diretamente a base de conhecimento e selecionar a parte que é de seu interesse. (Paley; Karp, 1997)
JOE (Java Ontology Editor)	Ferramenta para construção e visualização de ontologias. Proporciona gerenciamento do conhecimento em ambientes abertos, heterogêneos e com diversos usuários. As ontologias são visualizadas como um diagrama entidade-relacionamento, como o gerenciador de arquivos do <i>MS Windows</i> ou como uma estrutura em árvore. (Mahalingam; Huhns, 1997)
<i>APECKS (Adaptive Presentation Environment for Collaborative Knowledge Structuring)</i>	É um servidor de ontologias; permite trabalho cooperativo através da criação de ontologias pessoais pelos usuários. Estas ontologias podem ser comparadas com outras e é possível a discussão sobre as diferenças e similaridades entre elas. (Tennison; Shadbolt, 1998)
OilEd	É um editor de ontologias de código aberto; permite construir ontologias utilizando a linguagem OIL. Não é um ambiente completo para desenvolvimento de ontologias. Verificação da consistência e classificação automática da ontologia podem ser executadas pela ferramenta FaCT (Horrocks; Sattler; Tobies, 1999)
OntoEdit	É um ambiente gráfico para edição de ontologias; permite inspeção, navegação, codificação e alteração de ontologias. O modelo conceitual é armazenado usando um modelo de ontologia que pode ser mapeado em diferentes linguagens de representação. As ontologias são armazenadas em bancos relacionais e podem ser implementadas em XML, FLogic, RDF(S) e DAML+OIL. (Maedche et al., 2000)
OCM (<i>Ontological Constraints Manager</i>)	É uma ferramenta para verificar a consistência de ontologias em relação a axiomas ontológicos. É composta por duas ferramentas de edição que possibilitam verificar a ocorrência de conflitos. (Kalfoglou et al., 2001)
<i>Protege</i> 2000	É um ambiente interativo para projeto de ontologias, de código aberto, que oferece uma interface gráfica para edição de ontologias e uma arquitetura para a criação de ferramentas baseadas em conhecimento. A arquitetura é modulada e permite a inserção de novos recursos. (Noy; Ferguson; Musen, 2000)
<i>WebODE</i>	Ambiente para engenharia ontológica que fornece suporte à maioria das atividades de desenvolvimento de ontologias. Fornece uma API (<i>application programming interface</i>) que facilita a integração com outros sistemas, importa e exporta ontologias de linguagens de marcação. (Arpírez et al., 2001).
<i>WebOnto</i>	Ferramenta que possibilita a navegação, criação e edição de ontologias, representadas na linguagem de modelagem OCML. Permite o gerenciamento de ontologias por interface gráfica, inspeção de elementos, verificação da consistência da herança e trabalho cooperativo. Possui uma biblioteca com mais de cem ontologias. (Domingue, 2001)

FIGURA 2 - Ferramentas para construção de ontologias

As ferramentas para a construção de ontologias utilizam diferentes linguagens. A seção seguinte apresenta uma breve pesquisa sobre linguagens de construção de ontologias.

3.3 Linguagens disponíveis para a construção de ontologias

Alguns exemplos de linguagens para a construção de ontologias são apresentadas a seguir, na FIG. 3. A lista de linguagens aqui apresentada, assim como a lista de ferramentas mostrada na seção anterior, também não tem a pretensão de ser completa e exaustiva.

Linguagens	Breve descrição
Cycl	Linguagem formal cuja sintaxe é derivada lógica de primeira ordem. Expressa conhecimento, através de um vocabulário de termos (constantes semânticas, variáveis, números, seqüências de caracteres etc) os quais são combinados em expressões, sentenças e, finalmente, em bases de conhecimento. (Lenat; Guha, 1990)
<i>Flogic (Frame Logic)</i>	Integra <i>frames</i> e lógica de primeira ordem. Trata de uma forma declarativa os aspectos estruturais das linguagens baseadas em <i>frames</i> e orientadas a objeto (identificação de objetos, herança, tipos polimórficos, métodos de consulta, encapsulamento etc). Permite a representação de conceitos, taxonomias, relações binárias, funções, instâncias, axiomas e regras. (Kifer; Lausen; Wu, 1990)
LOOM	Descendente da família KL-ONE (<i>Knowledge Language One</i>) é baseado em lógica descritiva e regras de produção. Permite a representação de conceitos, taxonomias, relações n-árias, funções, axiomas e regras de produção. (Brill, 1993)
CARIN	Trata-se de uma combinação da <i>Datalog</i> (linguagem baseada em regras) e lógica descritiva ALN. Uma ontologia CARIN é construída por dois componentes terminológicos: um conjunto de conceitos com declarações de inclusão e um conjunto de regras que usam os conceitos. (Levy; Rousset, 1996)
GRAIL	É uma linguagem que especifica uma ontologia do domínio médico (GALEN).
Ontolíngua	Combina paradigmas das linguagens baseadas em <i>frames</i> e lógica de primeira ordem. Permite a representação de conceitos, taxonomias de conceitos, relações n-árias, funções, axiomas, instâncias e procedimentos. Sua alta expressividade causa problemas na construção de mecanismos de inferência. (Chaudhri et al., 1998)
OCML	Permite a especificação de funções, relações e classes, instâncias e regras. Utilizada em aplicações de gerenciamento do conhecimento, desenvolvimento de ontologias, comércio eletrônico e sistemas baseados em conhecimento. Aplicada em medicina, ciências sociais, memória corporativa, engenharia, portais da Web etc. (Domingue; Motta; Corcho, 1999; Chaudhri; KARP; Thomere, 1999).
<i>OML (Ontology Markup Language)</i>	É uma linguagem de especificação de ontologias baseada em lógica descritiva e grafos conceituais. Permite a representação de conceitos organizados em taxonomias, relações e axiomas em lógica de primeira ordem. (Kent, 1999)
RDF (<i>Resource Description Framework</i>)/RDFS (RDF Schema)	Desenvolvidos pelo W3 <i>Consortium</i> , tem por objetivo a representação de conhecimento através da idéia de redes semânticas. São linguagens não muito expressivas, permitindo apenas a representação de conceitos, taxonomias de conceitos e relações binárias. (Lassila; Swick, 1999)
NKRL	Desenvolvida para descrever modelos semânticos para documentos multimídia. Baseada em <i>frames</i> , permite definir conceitos organizados em taxonomias e relacionamentos entre eles. Possui uma sub-linguagem para a construção de argumentos complexos. (Bertino; Barbara; Zarri, 1999)
SHOE (<i>Simple HTML Ontology Extensions</i>)	Utiliza extensões ao HTML, adicionando marcações para inserir dados semânticos em páginas Web. As marcações podem ser: para a construção de

Metodologia	Breve descrição
	ontologias e para anotações em documentos da Web. Permite representar conceitos, taxonomias, relações, instâncias e regras de dedução. (Heflin; Hendler, 2000)
XOL	Trata-se de um pequeno subconjunto de primitivas do protocolo OKBC. É uma linguagem restrita, na qual apenas conceitos, taxonomias e relações binárias podem ser especificadas. Não possui mecanismos de inferência e foi projetada para o intercâmbio de ontologias no domínio da biomédica. (Karp, 1997)
OIL	É o precursor do DAML + OIL e a base para o projeto de linguagem para a Web Semântica. Combina várias primitivas de modelagem das linguagens baseadas em <i>frames</i> com a semântica formal e serviços de inferência da lógica descritiva. Pode verificar classificação e taxonomias de conceitos. (Fensel et al., 2001)
DAML + OIL	DAML + OIL é uma linguagem de marcação semântica para a Web que apresenta extensões a linguagens como o DAML (<i>DARPA Agent Markup Language</i>), RDF e RDFS, através de primitivas de modelagem baseadas em linguagens lógicas. Permite representar conceitos, taxonomias, relações binárias e instâncias. (Horrocks et al., 2001).

FIGURA 3 — Linguagens para construção de ontologias

A seção seguinte apresenta um roteiro para a construção de uma ontologia do domínio das referências bibliográficas, através da ferramenta OIEd baseada na linguagem OIL.

4 Roteiro para construção de uma ontologia utilizando a ferramenta OIEd

O papel de uma ontologia consiste em possibilitar expressividade em relação ao domínio do conhecimento que se pretende representar. Apresenta-se aqui um exemplo didático para a construção de uma ontologia bibliográfica, em que se podem buscar informações sobre o domínio.

O domínio da ontologia bibliográfica diz respeito à referências bibliográficas acadêmicas. Pode ser a base para determinar, por exemplo, uma política de ciência e tecnologia de um país ou região, baseando-se nas informações sobre autores, publicações, editores etc., contidas na ontologia. Entretanto, para que seja possível esse tipo de planejamento, são necessários estudos bibliométricos. A construção de uma ontologia real está além do escopo deste trabalho. Não se tem a pretensão de desenvolver uma ontologia definitiva para a organização de referências bibliográficas mas, apenas, um exemplo didático. O objetivo é apresentar um roteiro que ilustre a construção de ontologias. Utiliza-se um roteiro baseado nos estudos de Noy e Guinness (2001), cujas etapas são apresentadas em seguida.

4.1 Etapa 1: determinação do domínio e o escopo da ontologia

A ontologia tem com objetivo relacionar os tipos de dados que podem representar partes de uma referência bibliográfica. É usada para organizar os concertos do domínio e deve informar sobre as relações existentes entre referências bibliográficas, por exemplo: quais os trabalhos de um autor estão nas fontes consultadas, qual a relação de um trabalho com outros artigos presentes em outras bases etc.

Algumas consultas pertinentes que o conhecimento mapeado na

ontologia deve responder são apresentadas abaixo. As perguntas foram construídas a partir da consulta a trabalhos bibliométricos (Jacobs, 2001; Hawkins, 2001; Geisler, 1999). Uma ontologia sobre referências bibliográficas deve possibilitar resposta a questões como:

- *Sobre autores*: quantos artigos publicaram nos últimos cinco anos? Quantas são as citações referentes ao autor? Qual a titulação do autor?

- *Sobre as publicações*: quantas citações teve um artigo nos últimos X anos? Quais são os periódicos mais relevantes para a área? Quais os tipos de publicações mais produzidas?

- *Sobre a área de conhecimento*: quais são as palavras-chave da área? Que número médio de publicações é produzido na área por ano? Em que local se produz mais publicações?

Após determinar o escopo da ontologia que se pretende desenvolver, o próximo passo será a realização de consulta para se identificar ontologias desenvolvidas no domínio de interesse.

4.2 Etapa 2: pesquisa sobre ontologias existentes no domínio

Uma pesquisa em bibliotecas de ontologias disponíveis na Internet, além de fontes complementares, revela a existência ontologias para dados bibliográficos. Para a construção da ontologia bibliográfica são utilizados informações dessas seis ontologias, conforme FIG. 4:

Item	Endereço para obtenção de ontologias na Internet
(1)	< http://orlando.drc.com/daml/Ontology/Bibliographic/3.1/Bibliographic-ont.daml >
(2)	< http://www.daml.ri.cmu.edu/ont/homework/atlas-publications.daml >
(3)	< http://www.cs.yale.edu/~dvm/daml/bib-ont.daml >
(4)	< http://www.daml.ri.cmu.edu/ont/homework/cmu-ri-publications-ont.daml >
(5)	< http://lcweb.loc.gov/marc/umb/um07to10.html > (campos mais utilizados do MARC).
(6)	< http://orlando.drc.com/SemanticWeb/Topics/Ontology/Ontologies.htm > (<i>Dublin Core</i>).

FIGURA 4 - Endereços de algumas ontologias bibliográficas disponíveis na Internet

Com essas ontologias verifica-se quais termos são mais freqüentes no domínio, para que sejam aproveitados na nova ontologia. O MARC (padrão de registros catalográficos) e o *Dublin Core* (padrão de metadados) foram utilizados por sua relação direta com os dados bibliográficos.

Na seqüência, é criada uma lista abrangente de termos relacionados ao assunto, relações entre os conceitos ou outras propriedades. As classes e a hierarquia são definidas em etapa posterior e o objetivo agora se limita a uma *grande lista*.

Uma lista de possíveis termos da nova ontologia é composta por campos que aparecem nas ontologias já existentes, os quais devem ser agrupados em ordem alfabética. A lista deve conter o papel do conceito (classe, subclasse, propriedade etc). A partir dessa lista, são apresentados, na próxima seção, os conceitos e as hierarquias da nova ontologia.

4.3 Etapa 3: definição das classes, da hierarquia e das propriedades

Deve-se selecionar termos da lista citada na seção anterior que descrevem objetos e têm existência independente uns dos outros. Estes termos são as classes da ontologia e vão se tornar referências na hierarquia. Organizam-se as classes em uma taxonomia perguntando-se se, sendo instância de uma classe, o objeto será necessariamente uma instância para as outras classes. O raciocínio é se a classe A é uma superclasse da classe B, então cada instância de B é também uma instância de A (Noy e Guinness, 2001 p. 6). Por exemplo, todo vinho *Pinot Noir* é, necessariamente, um vinho tinto, e, dessa forma, a classe *Pinot Noir* é uma subclasse da classe vinhos tintos. Termos da lista podem também ser descartados.

Para montagem da ontologia, utiliza-se a ferramenta OILED (editor de ontologias). A FIG.5 mostra uma tela do editor onde se determinam as classes (ver lado esquerdo da tela).

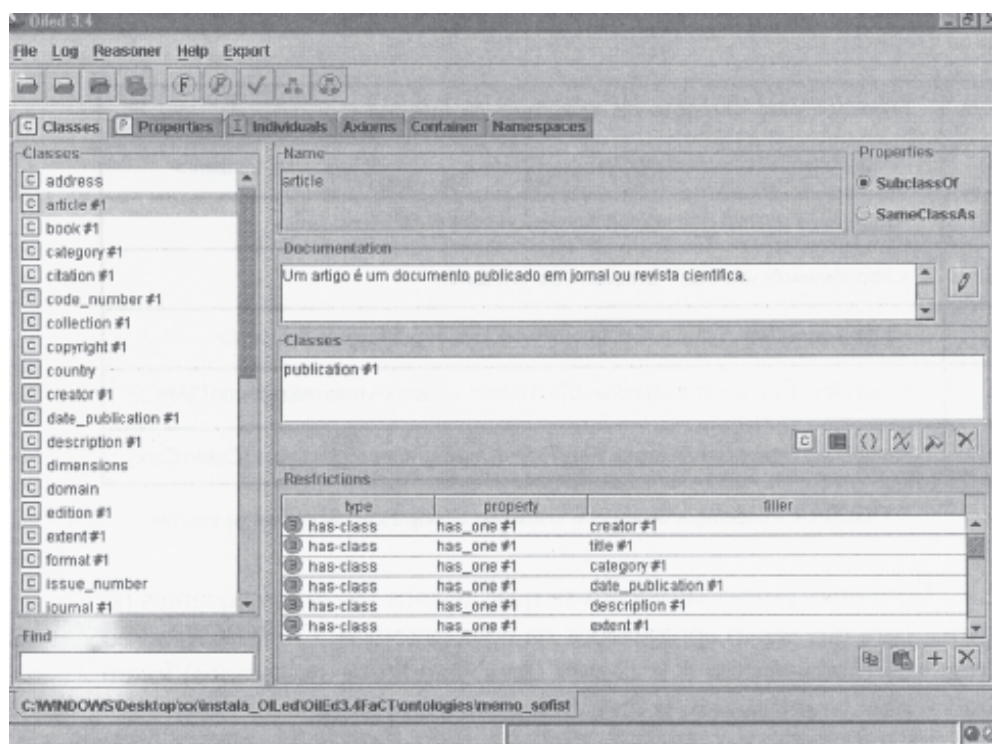


FIGURA 5(A) - Tela do editor de ontologias Oiled, mostrando as classes para ontologia

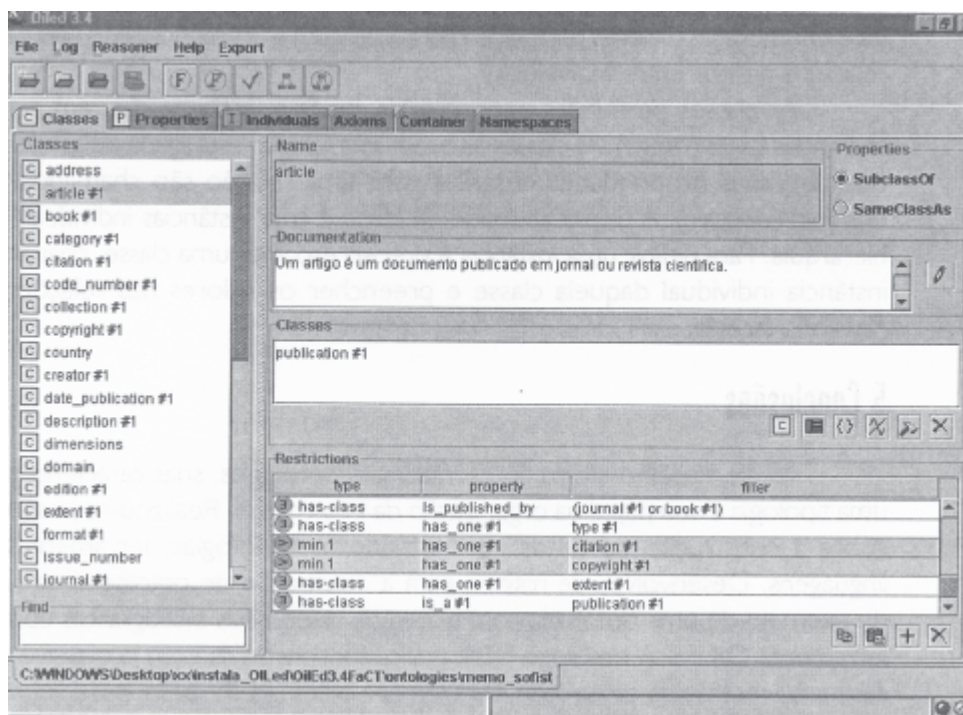


FIGURA 5 (B) - Tela do editor de ontologias, mostrando as propriedades da classe *article*

Com a lista citada na seção anterior, as classes são criadas através de um agrupamento semântico dos termos existentes. Por exemplo, são definidas classes que representam itens de acervo (*article*, *book* etc). Entretanto, apenas classes não possibilitam respostas às perguntas da seção 4.1. Termos que *sobram* na lista da seção 4.2 após a definição das classes podem ser propriedades das classes. Esses termos são, em geral, chamados de relações (*slots*).

Todas as subclasses de uma classe herdam a relação daquela classe. A relação deve se referir à classe mais geral que possui a propriedade. Na FIG. 5B as propriedades da classe *article* estão definidas no canto inferior direito da tela. Algumas relações definidas para a ontologia bibliográfica são: *authoring*, *correcting*, *has-domain*, *has-one*, *is-a*, *is-published-by*, *publish-one*, *publishing*.

4.4 Etapa 4: definição de restrições sobre as propriedades e determinação de instâncias

As *relações* tem características chamadas (não consensualmente) de *facetas* (*facets*), que descrevem o tipo dos valores, valores permitidos, a faixa de valores etc. Algumas facetas são: cardinalidade da relação, o tipo de valor da relação (*slot-value type*), o domínio e a faixa (*range*) de valores da relação.

A cardinalidade de uma relação define quantos valores ela deve ter. Na FIG. 5B a classe *article* tem duas relações de cardinalidade um: *citation* e *copyright*. Isso indica que um artigo tem sempre, no mínimo, uma *citação* e um *direito autorial*.

O tipo de valor da relação (*slot-value type*) descreve que tipos de valores podem ser utilizados, sendo os mais comuns: *strings*, números, booleanos,

listas enumeradas (valores permitidos para a relação) e do tipo instância (definem relacionamentos entre indivíduos).

As classes permitidas para relações do tipo instância são, em geral, chamadas faixas (*range*). As classes às quais uma relação está ligada ou as classes que têm suas propriedades descritas para uma relação são chamados de domínio (*domain*). A última atividade da etapa é criar instâncias individuais na hierarquia. Para definir uma instância é preciso escolher uma classe, criar uma instância individual daquela classe e preencher os valores das relações e possíveis facetas.

5 Conclusões

Neste trabalho foram apresentadas as ontologias, suas características, uma tipologia e seu papel na organização da informação. Realizou-se pesquisa sobre a construção de ontologias, listando metodologias, ferramentas e linguagens. Desenvolveu-se roteiro para a construção de ontologias através de exemplo de uma ontologia para a área de referências bibliográficas com a ferramenta OIEd. A ontologia bibliográfica foi construída com o objetivo de ilustrar o processo de construção. Para tornar a ontologia útil, pesquisas adicionais são necessárias.

De acordo com a tipologia apresentada na seção 2.2, a ontologia aqui apresentada é caracterizada como uma ontologia terminológica, ou seja, específica para termos no domínio; como uma ontologia de domínio, que descreve o vocabulário de um domínio; e, como uma ontologia semi-formal, expressa em linguagem artificial (OIL) definida formalmente.

O roteiro apresentado pode auxiliar a construção de pequenas ontologias. A ferramenta utilizada (OIEd) não é um ambiente completo de desenvolvimento. Diversas ferramentas e metodologias apresentadas podem ser utilizadas de acordo com a necessidade e complexidade requeridas. O OIEd gera automaticamente o código OIL (não apresentado aqui por limitações de espaço), que pode ser exportado para RDFS (escrita em RDF). Isso auxilia na representação da ontologia, visto que a RDF tem uma representação gráfica definida e padronizada.

A construção de ontologias pode facilitar o processo de recuperação da informação, ao caracterizar e criar taxonomias sobre conceitos de um domínio. Ao definir classes, relações, axiomas etc, possibilita-se a redução de diferenças terminológicas ou semânticas que prejudicam a eficiência da recuperação automática da informação.

Testar as ferramentas, linguagens e metodologias disponíveis para a construção de ontologias, estudando suas aplicações, pode auxiliar na escolha da melhor forma de construir ontologia e assim, organizar a informação em um domínio. Tal tarefa parece ser uma boa direção para novas pesquisas e espera-se realizá-la em trabalhos futuros.

A list of topics for construction of a bibliographic ontology using an automated tool

Recently, with the growing number of available data sources, it is increasingly difficult to select, acquire and combine data. Several published works have made use of ontologies to organize and characterize data. The use of terminological-based logics such as knowledge representation formalisms, allows the construction and implementation of ontologies in computers. This paper presents a brief survey of methodologies, tools and languages used to build ontologies and a list of topics to guide the construction of an ontology in the bibliographic field, with the use of OIL (Ontology Interchange Language) logic-based language. It also demonstrates how OILED can be used to build an ontology about bibliographic references.

Key-words: Ontology; Languages; Methodologies; Tools

Referências

AGUADO, G. et al. Ontogeneration: reusing domain and linguistic ontologies for Spanish text generation? In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 13th, ECAI'98. *Papers accepted to the workshop on applications of ontologies and problem solving methods*. Brighton, England, 1998. p. 23-28.

ALBERTAZZI, Liliana. Formal and material ontology. In: POLI, Roberto; SIMONS, Peter (Ed.). *Formal ontology*. Dordrecht: Kluwer, 1996. p. 199-232.

ALEXAKI, S. et al. Managing RDF metadata for community webs. In: INTERNATIONAL WORKSHOP ON THE WORLD WIDE WEB AND CONCEPTUAL MODELING, 2nd, 2000. p. 140-151. Disponível em: < <http://139.91.183.30:9090/RDF/publications/wcm2000.PDF> > . Acesso em: 11 out. 2002.

ARPIREZ, J. C. et al. Web ODE: a scalable workbench for ontological engineering. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE CAPTURE PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON KNOWLEDGE. Victoria, British Columbia, Canada, 2001.

BATEMAN, J. A. (1996). Using text structure and text planning to guide text summarization. Disponível em: < <http://www.ik.fhannover.de/fk/projekte/Dagstuhl/Abstract/Abstracts/Bateman/Bateman.html> > . Acesso em: 25 maio 2002.

BERTINO, E.; BARBARA, C.; ZARRI, G. P. A conceptual annotation approach to indexing in a web-based information system. In: INTERNATIONAL WORKSHOP ON ADVANCE ISSUES OF E-COMMERCE AND WEB-BASED INFORMATION SYSTEMS. Santa Clara, Califórnia, 1999.

BORGO, S. et al. Using a large linguistic ontology for Internet-based retrieval of object-oriented components. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING (SEKE 97), 9th. Madrid, Spain, 1997. *Proceedings..* Madrid, 1997.

BORSI, W. N. (1997). Construction of engineering ontologies. Thesis. Disponível em: < <http://www.utwente.nl/webdocs/mf/1/t000004.pdf> > . Acesso em: 21 abr. 2002.

BRILL, D. (1998). Loom reference manual. In: BACON, J.; HAYTON, R.; MOODY, K. *Middleware for digital libraries*. Disponível em: < <http://www.dlib.org/dlib/october98/bacon/10bacon.html> > . Acesso em: 15 nov. 2001.

CHANDRASEKARAN, B.; JOHNSON, T. R.; BENJAMINS, V. R. Ontologies: what are they? why do we need them? *IEEE Intelligent Systems*. v. 14, n. 1, p. 20-26, 1999.

CHAUDHRI, V. K. et al. (1998). Open knowledge base connectivity 2.0. Disponível em: < <http://www.ksl-svc.stanford.edu:5915/doc/release/okbc/okbc-spea/okbc-2-0-3.pdf> > . Acesso em: 9 jun. 2002.

CORCHO, O.; FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A. (2001). Ontoweb, technical roadmap V 1.0. Disponível em: < http://babage.dia.fi.upm.es/ontoweb/wp1/OntoRoadMap/documents/D11_v1_0.pdf > . Acesso em: 4 mar. 2002.

DOMINGUE, J. Tazebao and webonto: discussing, browsing and editing ontologies on the web. In: BANFF KNOWLEDGE ACQUISITION WORKSHOP 11 th., 1998. *Proceedings..* Banff, Alberta, Canada, 1998.

- DOMINGUE, J. et al. Supporting Ontology Driven Document Enrichment within Communities of Practice. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE CAPTURE. International conference on knowledge capture. *Proceedings*.. Victoria, British Columbia, Canada, 2001.
- DOMINGUE, J.; MOTTA, E.; CORCHO, O. (1999). Knowledge modeling in web onto and OCML: A user guide. Disponível em: < http://kmi.open.ac.uk/projects/webonto/user_guide.2.4.pdf >. Acesso em: 9 jun. 2002.
- EUZENAT, J. Corporative memory through cooperative creation of knowledge bases and hyper-documents. In: KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS WORKSHOP, 10th, 1996. *Proceedings*.. Disponível em: < <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/euzenat/euzenat96b.html> >. Acesso em: 28 set. 2001.
- FARQUHAR, A.; FIKES, R.; RICE, J. *The ontolingua server*. USA: a tool for collaborative ontology construction. Duluth: Academic Press, 1997. p. 707-727.
- FENSEL, D. et al. OIL: an ontology infrastructure for the semantic web. *IEEE Intelligent Systems*, Mai/Abr. 2001.
- FERNÁNDEZ-LÓPEZ et al. Building a chemical ontology using methontology and the ontology design environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, Jan./Feb. 1999.
- GANDON, F. Engineering an ontology for a multi-agents corporate memory system. In: INTERNATIONAL SYMPOSIUM ON THE MANAGEMENT OF INDUSTRIAL AND CORPORATE KNOWLEDGE. *Proceedings*.. 2001. p. 209-228. Disponível em: < <http://citeseer.nj.nec.com/gandon01engineering.html> >. Acesso em: 22 maio 2002.
- GEISLER, E. The metrics of technology evaluation: where we stand and where we should go from here. In: ANNUAL TECHNOLOGY TRANSFER SOCIETY MEETING. 24th, St. Pete Beach, FL, 1999. *Proceedings*.. 1999.
- GÓMEZ-PÉREZ, A. Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases. In: WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, 12th, Voyager Inn, Banff, Alberta, Canada, 1999.
- GUARINO, N.; GIARETTA, P. Ontologies and KBs, towards a terminological clarification. In: MARS, N. (Ed.). *Towards a very large knowledge bases: knowledge building and knowledge sharing*. [S.l.]: IOS Press, 1995. p. 25-32.
- GRUBER, T. (1996). What is an Ontology? Disponível em: < <http://www-ksl.stanford.edu/ksl/what-is-an-ontology.html> >. Acesso em: 14 set. 2002.
- GRÜNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING. Montreal, 1995.
- LEGER, A. et al. (2000). Ontology domain modeling support for multilingual services in e-Commerce: MKBEEM. *Presentation seminar ECAI2000* -Berlin. Disponível em: < <http://mkbeem.elibel.tm.fr/paper/ecai00-final.pdf> >. Acesso em: 22 dez. 2001.
- LEVY, A. Y.; ROUSSET, M. C. CARIN: A representation language integrating rules and description logics. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE. *Proceedings*.. Budapest, Hungary, 1996.
- LÓPEZ, M. F. (1999). *Over view of methodologies for building ontologies*. Disponível em: < <http://www.ontology.org/main/presentations/madrid/analysis.pdf> >. Acesso em: 31 jan. 2002.
- MARTIN, P. H.; EKLUND, P. (2001). *Large-scale cooperatively-built heterogeneous KBs*. In: ICCS'01, INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES., 9th. Disponível em: < <http://meganesia.int.gu.edu.au/~phmartin/WebKB/doc/papers/iccs01/> >. Acesso em: 6 set. 2002.
- MAHALINGAM, K.; HUHS, M. N. An ontology tool for query formulation in an agent-based context. In: IFCIS INTERNATIONAL CONFERENCE ON COOPERATIVE INFORMATION SYSTEMS (COOPIS '97), 2th, Kiawah Island, SC, 1997. *Proceedings*.. p. 170.
- MIZOGUCHI, R.; VANWELKENHUYSEN, J.; IKEDA, M. Task ontology for reuse of problem solving knowledge. In: ECAI'94 TOWARDS VERY LARGE KNOWLEDGE BASES. *Proceedings*.. Amsterdam: IOS Press, 1995. p. 46-59.
- NECHES, R. et al. Enabling technology for knowledge sharing. *AI Magazine*. v. 12, n. 3, Fall 1991.
- NOY, F. N.; GUINNESS, D. L. *Ontology development 101: a guide to create your first ontology*. Disponível em: < <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc> >. Acesso em: 4 maio 2001.
- NOY, N. E.; FERGERSON, R. W.; MUSEN, M. A. The knowledge model of Protege-2000: Combining interoperability and flexibility. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT (EKAW'2000), 2th, Juan-les-Pins, France, 2000. *Proceedings*..
- PALEY, J. L. S.; KARP, P. A. Generic Knowledge base Browser and Editor. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE; INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE CONFERENCE, 14th, 9th, Providence, Rhode Island, *Proceedings*.. AAAI Press/The MIT, 1997. p. 1.045-1.051.
- RECTOR et al. The Grail concept modeling language for medical terminology. *Artificial Intelligence in Medicine*, v. 9, p. 139-171, 1997.

- STAAB, S. et al. Knowledge processes and ontologies. *Intelligent systems*; Institute of Electrical and Electronics Engineers, IEEE, v. 16, n. 1, p. 26-34, Jan./Feb. 2001.
- SCHREIBER, A.; TERPSTRA, P.; SISYPHUS-VT (1995). *A common KADS solution*. Technical Report. Disponível em: < <http://www.swi.psy.uva.nl/projects/Kactus/Papers.html> > . Acesso em: 28 jun. 2002.
- SKUCE, D. CODE4: a Unified System for Managing Conceptual Knowledge. *International Journal of Human-Computer Studies*, n. 42, p. 413-451, 1995.
- SKUCE, D. (1996). IKARUS: *Intelligent knowledge acquisition and retrieval universal system*. Disponível em: < <http://www.csi.uottawa.ca/~kavanagh/ikarus/ikarus4.html> > . Acesso em: 4 out. 2002.
- SOWA, J. E. (1999). *Building, sharing and merging ontologies*. Tutorial < <http://users.bestweb.net/~sowa/ontology/ontoshar.htm> > .
- SWARTOUT, B. et al. Toward distributed use of large-scale ontologies. In: AAAI97 SPRING SYMPOSIUM SERIES WORKSHOP ON ONTOLOGICAL ENGINEERING. *Proceedings...* AAAI Press, 1997. p. 138-148.
- TENNISON, J.; SHADBOLT, N. R. APECKS: a tool to support living ontologies. In: BANFF KNOWLEDGE ACQUISITION WORKSHOP, 11th, Banff, Alberta, Canada. *Proceedings...* 1998.
- USCKOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.
- USCHOLD, M.; KING, M. Building ontologies: towards a unified methodology. In: ANNUAL CONFERENCE OF THE BRITISH COMPUTER SOCIETY SPECIALIST GROUP ON EXPERT SYSTEMS, 16th, Cambridge, UK. *Proceedings...* Cambridge, 1996.
- VAN-HELIJST, G.; SCHREIBER, A. T; WELINGA, B. J. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, v. 46, Issue 2-3, p. 183-192, Feb./Mar. 1997.
- VÁZQUEZ, E.; VALERA, F.; BELLIDO, L. (2001). *Modelado de servicios complejos en una plataforma de intermediación para comercio electrónico*. Disponível em: < <http://www.telecom.ece.ntua.gr/smartec/documentation/Publications/smartec-jitel2001.pdf> > . Acesso em: 14 ago. 2002.