

Maurício Barcellos Almeida

**INTER-OPERABILIDADE ENTRE FONTES
HETEROGÊNEAS: UM META-MODELO BASEADO
EM ONTOLOGIAS**

Belo Horizonte

2002

Maurício Barcellos Almeida

**INTER-OPERABILIDADE ENTRE FONTES
HETEROGÊNEAS: UM META-MODELO BASEADO
EM ONTOLOGIAS**

Dissertação apresentada ao Curso de mestrado da Escola de
Ciência da Informação da Universidade Federal de Minas
Gerais, como requisito parcial à obtenção do título de mestre
em Ciência da Informação.

Área de concentração: Informação gerencial e tecnológica

Orientador: Prof. Dr. Marcello Peixoto Bax

Belo Horizonte

2002

À meus pais,
Para **Lucy** (*in memoriam*),
que nos deixou nesse ano de 2002 ...

“Mesmo na distância e no silêncio, a
ternura nos mantém unidos”.

Para **Corette** (*in memoriam*),

“Um dia as lágrimas secarão, mas
nunca esqueceremos os caminhos da
vida que você nos indicou”.

Agradecimentos

Aos Professores e Funcionários da ECI - UFMG

Sra. Goreth, por sua presteza

Profa. Beatriz Cendón, por sua disponibilidade

Prof. Madalena Naves, por sua gentileza

Prof. Marcello Bax, por sua contribuição

Aos professores da PUCMINAS

Prof. Humberto Torres Marques Neto,

pelas longas conversas sobre informação e computação

Em especial, ao *prof. Manoel Palhares*, pela amizade, apoio e confiança

“Sócrates disse certa vez que ele atuava como uma ‘parteira’ para seus ouvintes, pois cuidava para que eles refletissem melhor a respeito do que já sabiam e se tornassem conscientes disso. Se nós apenas enxergássemos o que já sabemos, isto é, no uso de certos conceitos e palavras que são tão sutis , ficaríamos atônitos com os tesouros contidos em nosso conhecimento.”

Immanuel Kant (1724-1804)

SUMÁRIO

LISTA DE FIGURAS	7
CAPÍTULO 1 – INTRODUÇÃO	9
CAPÍTULO 2 – INTER-OPERABILIDADE EM AMBIENTES DIGITAIS HETEROGÊNEOS	12
2.1 Introdução	12
2.2 Características e complexidade dos ambientes heterogêneos.....	12
2.3 De que trata a inter-operabilidade?	17
2.4 Que nível de integração é necessário?	19
2.5 Classificação de elementos em modelos de inter-operabilidade.....	21
2.6 Pesquisas sobre inter-operabilidade	23
2.7 Considerações finais.....	24
CAPÍTULO 3 – ONTOLOGIAS	25
3.1 Sobre representação do conhecimento	25
3.2 Sobre as Ontologias.....	27
3.2.1 Relação entre ontologias, lógica e organização do conhecimento	27
3.2.2 Definições para ontologia.....	30
3.2.3 Características de uma ontologia.....	33
3.2.4 Tipos de Ontologias	35
3.2.5 O papel das ontologias nos sistemas atuais	37
3.2.5 Ontologias disponíveis na Internet.....	37
3.3 Breve pesquisa sobre a construção de ontologias	39
3.3.1 Metodologias para a construção de ontologias.....	40
3.3.2 Ferramentas e linguagens para construção de ontologias.....	40
3.4 Considerações finais.....	41
CAPÍTULO 4 – ONTOLOGIAS E INTER-OPERABILIDADE	42
4.1 Introdução	42
4.2 Aplicações baseadas em ontologias	42
4.3 Trabalhos relacionados	44
4.3.1 Aplicações baseadas em ontologias específicas para a integração.....	44
4.3.2 <i>TSIMMIS</i>	45
4.3.3 <i>SIMS</i>	46
4.3.4 <i>COIN</i>	47
4.3.5 <i>OBSERVER</i>	48
4.3.6 <i>INFOSLEUTH</i>	49
4.3.7 <i>KRAFT</i>	50
4.3.8 <i>PICSEL</i>	52
4.3.9 <i>DWQ</i>	53
4.3.10 <i>ONTOBROKER</i>	54
4.3.11 <i>SHOE</i>	54
4.3.12 <i>BUSTER</i>	55
4.4 Quadro sinótico	56
CAPÍTULO 5 – A PROPOSTA PARA UM META-MODELO DE INTEGRAÇÃO.....	58
5.1 Introdução	58
5.2 Modelos e meta-modelos	60
5.3 Uma proposta preliminar para uma ontologia bibliográfica.....	61
5.3.1 Introdução	61
5.3.2 Domínio e o escopo da ontologia.....	62

5.3.3	Ontologias existentes.....	64
5.3.4	Definição das classes, da hierarquia e das propriedades	65
5.3.5	Definição das restrições sobre as propriedades	66
5.3.6	Considerações finais.....	67
5.4	O meta-modelo para integração.....	68
5.4.1	A arquitetura de integração do meta-modelo	68
5.4.2	Descrição informal do meta-modelo	72
5.4.3	Representação do meta-modelo	77
CAPÍTULO 6 – DISCUSSÃO.....		79
6.2	O contexto de estudo do nível semântico.....	79
6.3	A solução adotada	80
6.4	Questões pendentes e trabalhos complementares	82
6.5	Propostas para aplicação do presente trabalho	83
CAPITULO 7 – CONCLUSÕES.....		85
BIBLIOGRAFIA.....		87
Referências bibliográficas		87
Bibliografia complementar.....		97
Anexos		103
Anexo A –	Algumas definições e símbolos da lógica descritiva	104
Anexo B –	O processo de integração baseado em lógica descritiva	111
Anexo C –	Sobre o OIL, RDFe o RDFS	122
Anexo D –	Sintaxe do meta-modelo em OIL e RDFS	125
Anexo E –	Metodologias, ferramentas, linguagens para construção e projetos que fazem uso de ontologias.....	129
Anexo F –	Lista de classes, propriedades e comentários de algumas ontologias sobre registros bibliográficos.....	135
Anexo G –	Arquiteturas de diversos sistemas para integração	143

LISTA DE FIGURAS

Figura 1 – Tipos de heterogeneidade verificadas no mundo	12
Figura 2 – Extração de dados de várias fontes e ontologia.....	13
Figura 3 – Heterogeneidade em ambientes informacionais de redes	14
Figura 4 – Heterogeneidade de sistemas de informação	15
Figura 5 – Diferentes expressividades na codificação XML.....	16
Figura 6 – Diferentes línguas (inglês e holandês) com o mesmo sentido.....	16
Figura 7 – Classificação para a inter-operabilidade - <i>UK Interoperability Focus</i>	18
Figura 8 – Esquema conceitual como núcleo de um	29
sistema integrado de dados (SOWA, 1999b)	29
Figura 9 – Formas de aplicação de ontologias (USCHOLD; GRÜNINGER, 1996).....	43
Figura 10 – Quadro resumo de projetos que utilizam ontologias.....	57
Figura 11 – Tela do editor de ontologias OILed mostrando.....	65
as classes para a ontologia bibliográfica.....	65
Figura 12 – Tela do editor de ontologias OILed mostrando as	67
restrições de cardinalidade um.....	67
Figura 13 – Arquitetura da camada conceitual	69
Figura 14 – Níveis de modelagem.....	70
Figura 15 – Tela do OILed, editor de ontologias utilizado para construção do meta-modelo	72
Figura 16 – Tela do OILed, com representação das hierarquias definidas entre as classes	77
Figura A1 – Exemplos de construtores para a lógica descritiva.....	104
Figura A2 – Exemplos de construtores para a lógica descritiva.....	104
Figura A3 – Domínio da interpretação	105
Figura A4 – Definições intensionais e extensionais	106
Figura B1 – Modelo para as assertiva intra-modelo MF_2	118
Figura B2 – Modelo para a assertiva intra-modelo MF_3	119
Figura B3 – Visão reconciliada	119
Figura B4 – Ilustração do significado de “inferência”	120
Figura E1 – Metodologias para construção de ontologias sem modelos,.....	129
reutilização ou importação de outras ontologias.....	129
Figura E2 – Metodologias para construção de ontologias em grupo.....	129
Figura E3 – Metodologias para aprendizado da estrutura de ontologias.....	129
Figura E4 – Metodologias para integração de ontologias	130
Figura E5 – Métodos para avaliação de ontologias.....	130
Figura E6 – Projetos relacionados à gestão do conhecimento.....	131

Figura E7 – Projetos relacionados à comércio eletrônico	132
Figura E8 – Projetos relacionados ao processamento de linguagens naturais.....	132
Figura E9 – Projetos relacionados à recuperação da informação na Web.....	133
Figura E10 – Projetos relacionados educação	134
Figura F1 – Lista de campos relacionados por ordem alfabética	137
Figura F2 – Agrupamento semântico dos termos e definição de classes / subclasses.....	138
Figura F3 – Classes, seus <i>slots</i> e <i>facets</i>	142

CAPÍTULO 1 – INTRODUÇÃO

A explosão da disseminação da informação iniciada a partir do grande avanço das técnicas computacionais na década de 90 tem trazido muitos benefícios mas restam muitos problemas. Nos dias de hoje, o sucesso das instituições, e mesmo das nações, está intimamente relacionado ao uso que fazem das informações que podem manipular. Entretanto, para execução das tarefas necessárias às empresas, instituições e outros setores da sociedade, as pessoas recebem uma sobrecarga de informações além de sua capacidade. Muitas vezes não conseguem acessar as informações que seriam relevantes para a solução de seus problemas.

A evolução dos equipamentos e dos métodos de construção de aplicativos, a emergência da Internet e das redes privadas, o enorme avanço nas telecomunicações, o aumento constante na capacidade de processamento e armazenamento, nem sempre têm sido acompanhados por evolução correspondente das técnicas para acesso a informações.

Vários desses problemas dizem respeito à dificuldade na seleção, aquisição e combinação de informações nos repositórios digitais de dados. Hoje são possíveis grandes sistemas distribuídos, situados em diferentes localidades, visto a facilidade de acesso a redes, o que, paradoxalmente, tem dificultado em muito a recuperação da informação desejada. Um problema relevante diz respeito à heterogeneidade das diversas fontes de dados disponíveis no labirinto que se tornaram as redes digitais.

A dificuldade em acessar fontes de dados em função de sua heterogeneidade é a motivação para este trabalho. O processo de se buscar a inter-operabilidade entre fontes de dados diversas e heterogêneas constitui um problema complexo e que demanda diferentes abordagens para sua solução. A literatura na área é extensa. Em muitos casos, abordagens combinadas são necessárias para que os resultados sejam satisfatórios. Um trabalho de sistematização das iniciativas de pesquisa na área é importante para se possa conhecer as diversas formas de abordar do problema.

Os ambientes heterogêneos estão em toda parte, em organizações, universidades, instituições e em espaços como a Internet. A Ciência da Informação, como campo dedicado às pesquisas científicas voltadas para os problemas da efetiva comunicação do conhecimento e seus registros entre os seres humanos (SARACEVIC, 1996), encontra, nesse contexto, boas oportunidades para discussão das questões de inter-operabilidade.

Neste trabalho, estudam-se fontes heterogêneas de dados, suas relações e características, suas possíveis formas de acesso e integração, dentre outros aspectos. Desenvolve-

se aqui, segundo classificação proposta por Lakatos e Marconi (1991), uma pesquisa aplicada, quanto a sua natureza; qualitativa, quanto à forma de abordar o problema; exploratória, quanto a seus objetivos; a qual procura contribuir com pesquisa bibliográfica em fontes secundárias sobre soluções para a inter-operabilidade entre dados heterogêneos em ambientes digitais. Além disso, espera-se apresentar um modelo para integração de fontes. Para tal abordagem será necessário estudar a heterogeneidade dos dados sob diferentes aspectos: diversidade de formatos, tecnologias, tipos de usuários, ambientes e plataformas, dentre outras variáveis.

Os objetivos específicos a que esse trabalho se propõe dizem respeito ao estudo das principais características de fontes de dados heterogêneos e o entendimento das questões de representação da informação que permeiam os problemas de integração de fontes de dados. Em representação da informação estuda-se como ferramenta a lógica descritiva¹. Além disso, estuda-se a necessidade de inter-operabilidade em ambientes heterogêneos representados por grandes sistemas de informação, ambientes de rede (privadas ou Internet), bibliotecas digitais, etc. Pretende-se ainda apresentar as principais iniciativas e trabalhos relacionados na área além de tentativas de se obter classificação para os componentes de um projeto de inter-operabilidade. A contribuição da revisão bibliográfica é proporcionar ao leitor uma visão geral do “estado da arte” no assunto em questão.

Um problema dos sistemas bibliográficos tradicionais é que estes consideram que os objetos relevantes para busca são apenas livros nas prateleiras de bibliotecas, os quais tem atributos como autor, título, data, editor, etc. Novos sistemas baseados em ontologias² propõem novos objetos, como pessoas, organizações, conferências, locais, eventos, etc, cada um com seus atributos próprios. Os usuários podem navegar por novas informações e fazer consultas sobre esses novos objetos e não apenas sobre livros, num processo interativo.

O objetivo geral deste trabalho diz respeito à construção de um meta-modelo que possibilite integração de fontes heterogêneas de referências bibliográficas. O meta-modelo procura prover inter-operabilidade enfatizando aspectos semânticos, com a utilização de ontologias. A viabilidade do uso de ontologias como solução para inter-operabilidade será avaliada ao longo do trabalho. A idéia do meta-modelo pode ser utilizada em situações diversas. Em particular, espera-se integrar o modelo resultante ao escopo do projeto Sistema de Agentes para Bibliotecas On-Line (SABiO) (BAX, 1997), em andamento na Escola de Ciência da

¹ Lógicas descritivas são linguagens de representação do conhecimento construídas especialmente para expressar conhecimento através de conceitos e hierarquias.

² Uma especificação formal e explícita de como representar os objetos, conceitos e outras entidades que se assume existir em alguma área de interesse e os relacionamentos que elas mantêm entre si. O significado dos termos será discutido no Capítulo 3.

Informação da UFMG. Essa última tarefa representa a contribuição prática do trabalho. O SABiO consiste de um sistema composto por agentes³ que buscam extrair, combinar, integrar e apresentar dados provenientes de fontes bibliográficas distintas, em Ciência da Informação e em português.

A contribuição global que se busca, em última instância, é cooperar na construção do conhecimento de uma área de pesquisa que objetiva melhorar o acesso, a disseminação e a disponibilidade de fontes para uso mais eficiente por profissionais da informação. Acredita-se que esta contribuição seja possível e importante para as pesquisas da Ciência da Informação.

Um primeiro passo no sentido de entender o que se procura tratar aqui é conhecer as características da heterogeneidade que se apresenta nos ambientes digitais. O Capítulo 2 se propõe a esta tarefa, estudando que tipo de integração e que nível de inter-operabilidade está se buscando e apresentando propostas para classificação de elementos em ambientes heterogêneos. O Capítulo 3 apresenta as ontologias, discute conceitos, características, tipos, técnicas para manipulação e construção, e, define o que significa o termo “ontologia” no escopo deste trabalho. Além disso, apresenta uma breve discussão sobre a importância das ontologias e sobre como o estudo do problema abrange diversas comunidades de pesquisa. O Capítulo 4 aborda a utilização de ontologias em projetos de inter-operabilidade, apresentando aplicações e trabalhos relacionados. No Capítulo 5 constrói-se uma proposta preliminar para uma ontologia bibliográfica que poderá ser utilizada como modelo para o SABiO. Além disso, constrói-se também um meta-modelo, baseado em ontologias, que será aplicado ao processo de integração de fontes do projeto SABiO. O Capítulo 5 traz ainda referências ao processo de integração cujo formalismo matemático é apresentado nos Anexos. O Capítulo 6 discute questões sobre o uso do modelo, sua viabilidade, e, de forma breve, tece algumas considerações sobre a possibilidade de implementação da lógica descritiva (utilizada para a integração). O Capítulo 7 apresenta as conclusões, as lições aprendidas e indica diretrizes para trabalhos futuros. Nos Anexos apresentam-se definições e simbologia da lógica descritiva, o formalismo matemático baseado em lógica descritiva que representa a integração de fontes heterogêneas, conceitos necessários ao entendimento do meta-modelo, comentários sobre RDF⁴ e OIL⁵, além da sintaxe do meta-modelo em RDFS e OIL. Apresentam-se informações sobre ferramentas para construção de ontologias e projetos relacionados, além de tabelas auxiliares que complementam os anexos.

³ Artefato de *software* que executa algum tipo de processo automático, no qual uma entidade pode se comunicar com outras similares, para executar tarefas e representar uma ou mais pessoas.

⁴ *Resource Description Framework*, padrão de metadados para meios digitais. Para maiores detalhes, consultar Anexo C.

⁵ *Ontology Interchange Language*, linguagem especializada na construção de ontologias. Para maiores detalhes, consultar Anexo C.

CAPÍTULO 2 – INTER-OPERABILIDADE EM AMBIENTES DIGITAIS HETEROGÊNEOS

2.1 Introdução

Ao se falar de heterogeneidade em ambientes digitais, pode-se estar lidando com diferentes perspectivas. O assunto é complexo e não diz respeito apenas a diferentes formatos de dados, como se pode imaginar numa abordagem superficial. O formato que representa os dados faz parte do problema, mas corresponde apenas a uma parcela do que é necessário tratar.

Apresenta-se nessa seção uma visão geral das características dos ambientes informacionais atuais, enfatizando sua heterogeneidade. Além disso são apresentadas questões relativas a como se buscar a inter-operabilidade. Para tanto, discute-se que grau de integração é desejado e do que realmente se está tratando ao se buscar inter-operabilidade. Apresentam-se também, de forma breve, propostas para classificação de elementos presentes em projetos que buscam integração. Espera-se que essas propostas possam auxiliar na compreensão do assunto.

2.2 Características e complexidade dos ambientes heterogêneos

Muitos estudos que objetivam a integração de fontes de heterogêneas não levam em conta a diversidade de tipos de heterogeneidade que ocorrem nos ambientes onde sistemas interagem. Uma observação mais atenta pode identificar diversos níveis em que a heterogeneidade se manifesta.

Ouksel (1992)⁶ citado por Ouksel e Sheth (1999), apresenta um esquema genérico para classificação da heterogeneidade, apresentado na Figura 1:

Mundo Social: crenças, expectativas, compromissos, contratos, leis, cultura ...
Pragmática: intenções, comunicação, conversações, negociações ...
Semântica: significados, proposições, validade, honestidade, significado, definições ...
Sintaxe: estrutura formal, linguagem, lógica, dados, registros, deduções, softwares, arquivos ...

Figura 1 – Tipos de heterogeneidade verificadas no mundo

Do ponto de vista estritamente tecnológico, o estudo da heterogeneidade também é bem amplo. Modelos, plataformas, *softwares*, formatos de arquivos eletrônicos, tipos de bancos de dados, utilitários de armazenamento, dentre outros; compõem um ambiente altamente

⁶ OUKSEL, A. Semantic mechanisms for cooperation in heterogeneous database systems. *Journal of Intelligent Information Systems*, v. 1, n. 3, p. 151-183, 1992.

complexo e dinâmico. Os dados podem estar em formatos texto, vídeo ou áudio, com graus variáveis de estrutura (dados estruturados, como os bancos de dados relacionais; dados semi-estruturados, como os *e-mails* e as páginas *HTML*⁷; dados não-estruturados, como os textos). Os documentos, em forma de arquivos digitais, representam a parte do conhecimento armazenado em computadores. A maioria desses documentos tem formato adequado para leitura humana, e o conhecimento que contêm, nem sempre pode ser acessado facilmente por *softwares* (GAO; STERLING, 2000). Dessa forma, na maioria dos casos, o processo de integração automatizada não é simples.

A Figura 2 apresenta um esquema genérico onde dados estruturados, semi-estruturados e sem estrutura alimentam uma hierarquia, representada pela ontologia⁸, onde são feitas as consultas do usuário.

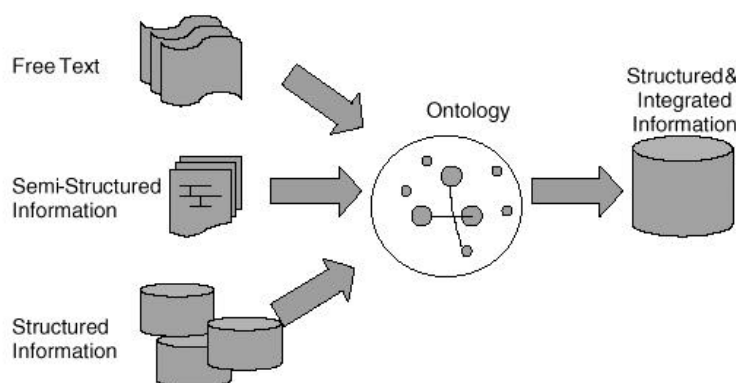


Figura 2 – Extração de dados de várias fontes e ontologia

Ainda com ênfase no aspecto tecnológico, a integração de sistemas corporativos é um problema usual nas áreas de vendas, engenharia, produção, contabilidade, finanças, administração de materiais, logística, etc. (WATSON, 2000). O problema se intensifica quando se trata da integração dos sistemas de diferentes organizações, de onde resulta um grau de incompatibilidade ainda maior. Algumas razões para tal incompatibilidade são o uso de várias linguagens de programação, de linguagens já em desuso⁹, de sistemas operacionais diversos¹⁰ e em padrões de conectividade para rede¹¹.

Weihai (2002) também apresenta estatísticas sobre dados heterogêneos de sistemas

⁷ *Hypertext markup language*, linguagem de marcação, atual padrão em uso na Internet.

⁸ O assunto será estudado em detalhe no Capítulo 3 e a definição do termo, mais especificamente, na Seção 3.2.2.

⁹ Por exemplo o *COBOL-Common Business Oriented Language*; uma linguagem, já em desuso, utilizada em aplicativos comerciais.

¹⁰ Por exemplo o UNIX, MacOS, NT, Windows, Windows VE, Palm OS, etc.

¹¹ Por exemplo, Ethernet, ATM, IP, Appletalk, USB, etc.

legados¹² e a necessidade de integrá-los. Segundo o autor, existem ainda no mundo “trinta bilhões de linhas de código COBOL, independentemente desenvolvidos, mantidos e gerenciados, na forma de sistemas legados, os quais não podem, simplesmente, ser descartados”.

A competitividade das organizações, no ambiente de negócios atual, depende da qualidade do acesso à sua *base de conhecimentos*¹³. Essa tarefa tem se mostrado árdua. A Internet, *intranets*, redes de alta velocidade e infra-estruturas de computação distribuída continuam a ganhar em popularidade como meios eficientes de comunicação. Em tais ambientes, caracterizados por fontes de dados altamente distribuídas, o acesso à informação relevante torna-se cada vez mais complexo.

Assim, a heterogeneidade é, sobretudo, considerada em ambientes informacionais de rede que são a principal forma de acesso à dados nas organizações e instituições. O estudo da heterogeneidade em redes, aí incluída a Internet, constitui um importante tópico de pesquisa (BALDONADO; COUSINS, 1997). Desse ponto de vista, os principais tipos de heterogeneidades estão relacionados a:

Usuários: <i>Estudos etnográficos bem conhecidos¹⁴ estabeleceram que populações de usuários exibem grande heterogeneidade. Podem-se distinguir cinco importantes dimensões deste tipo de heterogeneidade: faixa de atividades, experiência, estilo, localização geográfica e ferramentas disponíveis. As atividades empreendidas pelos usuários individuais podem incluir escrita e publicação da informação; coleta, organização e análise de recursos; comunicação e trabalho em grupo; etc.</i>
Repositórios de informação: <i>Exemplos de repositórios incluem as coleções tradicionais de bibliotecas, imagens digitais, arquivo de e-mail, vídeo, livros on-line e catálogos de citações de artigos científicos (contendo apenas meta-informações). Estes repositórios diferem uns dos outros em tipo, em gênero (livros x filmes), modalidade (imagens x textos), em assunto (entretenimento x ciência) e em esquemas empregados para descrevê-los (conhecidos como esquemas de catalogação na biblioteconomia).</i>
Serviços de informação: <i>Espera-se que uma grande diversidade de serviços de informação independentes vá emergir no ambiente de redes, comosumarização, tradução, arquivamento, detecção de cópias, publicação, assistência à busca de informações e distribuição de documentos.</i>

Figura 3 – Heterogeneidade em ambientes informacionais de redes

Fazendo um paralelo entre as Figuras 1 e 3, pode-se corresponder os “usuários” à

¹² Geralmente desenvolvidos no passado, estes sistemas foram construídos em linguagens e padrões já em desuso, para os quais, muitas vezes, não se encontram mais profissionais qualificados ou os próprios autores para manutenção.

¹³ Conhecimento expresso com a utilização de alguma linguagem formal de representação do conhecimento. Uma base de conhecimento faz parte de um sistema baseado em conhecimento.

¹⁴ Baldonado e Cousins (1997) apresentam duas referências para os estudos etnográficos citados.

pragmática. A heterogeneidade do mundo social (primeira linha da Figura 1) trata do assunto de maior representatividade dentre os citados, pois se procura integrar comunidades de pessoas, o que na maioria das vezes não é simples (e muitas vezes impossível). Nesse contexto, a pragmática corresponde ao ramo da semiótica que trata dos sinais ou expressões lingüísticas e aqueles que os usam. Aos “repositórios de informação” nas redes digitais, pode-se corresponder à sintaxe, pois a inter-operabilidade necessária para a operação conjunta de bancos de dados muitas vezes depende exclusivamente da sintaxe. O aplicativo não tem a preocupação de verificar se os termos conjugados fazem ou não sentido. “Serviços de informação” podem estar ligados à semântica, de onde se extrai o “sentido das coisas”.

Dentre os tipos citados, a integração pragmática representa um desafio, por levar em conta aspectos subjetivos da natureza humana. A busca de integração relativa à pragmática não será abordada nesse trabalho. Dentre os tipos restantes, o de maior complexidade e importância é a heterogeneidade semântica. Em relação a sintaxe, será necessária uma escala de padronização que parece improvável.

Sheth (1998)¹⁵ citado por Oukse e Sheth (1999), apresenta outra classificação para tipos de heterogeneidade, localizando o problema no contexto dos sistemas de informação. Esta figura já não apresenta, como as anteriores, a indicação à pragmática e se inicia na semântica, enfatizando em seguida, os sistemas:

<i>Heterogeneidade da informação</i> <i>Heterogeneidade semântica</i> <i>Heterogeneidade estrutural, representacional/esquemática</i> <i>Heterogeneidade de formatos e sintática</i>
<i>Heterogeneidade de sistemas</i> <i>Heterogeneidade de sistemas de informação</i> <i>Sistemas de gerenciamento de repositórios de mídia digital</i> <i>Sistemas de gerenciamento de banco de dados (SGBDs-Sistemas Gerenciadores de banco de dados, modelos de dados, capacidades de sistemas como controle de concorrência e recuperação)</i> <i>Heterogeneidade de plataformas</i> <i>Sistemas operacionais (heterogeneidade de sistemas de arquivos, tipos de arquivos, operações, suporte à transações)</i> <i>Hardware e sistemas (heterogeneidade de conjunto de instruções, representação de dados e codificação)</i>

Figura 4 – Heterogeneidade de sistemas de informação

Muitas diferenças sintáticas podem ser resolvidas com o uso do XML-Extended

¹⁵ SHETH, A. *Changing focus on interoperability in information systems: from systems, syntax, structured to semantics, in interconnecting geographic operation systems*. Norwell, MA M.F.Goodchild, M.J. Egenhofer, R.Fegeas, C.A.Kottman Kluwer Pub., 1998.

Markup Language, padrão que facilita o intercâmbio de dados ao uniformizar sua apresentação em diferentes aplicativos. Entretanto, o XML não pode solucionar todo tipo de heterogeneidade semântica. Na Figura 5, os problemas semânticos são de difícil solução pelo XML, pois são relativos às diferentes representações e suas respectivas capacidades de expressão (OMELAYENKO, 2001):

<pre><price> 699 USD </price></pre>	<pre><price> <amount> 750 </amount> <currency> DM </currency> </price> <price> <amount> 699 </amount> <currency> USD </currency> </price></pre>
(a)	(b)

Figura 5 – Diferentes expressividades na codificação XML

Na Figura 5, uma representação é semanticamente mais rica que a outra. Além disso, pode-se atribuir sinônimos na mesma língua ou em línguas diferentes para *tags*¹⁶ de mesmo significado, conforme Figura 3:

<pre><price> <sum> 990 </sum> <currency> DM </currency> <payment>transfer</payment> </price></pre>	<pre><prijs> <bedrag> 990 </bedrag> <valuta> DM </valuta> <afbetaling> transfer </afbetling> </prijs></pre>
(a)	(b)

Figura 6 – Diferentes línguas (inglês e holandês) com o mesmo sentido

Na Figura 6, em (a) os *tags* estão em inglês e em (b) estão em holandês, mas são semanticamente iguais. Mesmo utilizando uma única língua pode-se ter diversos sinônimos para o nome dos *tags*. O sistema de integração tem que reconhecer os nomes de *tags* sinônimos e traduzi-los. Existe ainda a possibilidade de uma organização definir entidades com semântica diferente.

Soluções para os desafios apresentados requerem abordagens diferenciadas das questões de heterogeneidade. Torna-se importante esclarecer que tipo de inter-operabilidade se procura, quais os problemas se pretende resolver e quais níveis de integração são necessários no contexto em estudo. Tal assunto será abordado na Seção 2.4 após uma breve introdução sobre de que trata a inter-operabilidade, apresentada na Seção 2.3 a seguir.

Conforme citado, estudam-se aqui as possíveis diferenças conceituais ou terminológicas entre fontes de informação. Para se alcançarem os objetivos propostos, é

¹⁶ Marcação caracterizada pelos sinais < e />.

importante estudar quais são os níveis de integração necessários. O valor agregado à processos que utilizam diferentes fontes de dados é gerado através de acesso amplo e consistente à informação.

2.3 De que trata a inter-operabilidade?

Várias possibilidades são encontradas na literatura envolvendo o termo inter-operabilidade. Bergamaschi e Sartori (1997) destacam que a extração de informações de fontes diversas de dados e sua integração podem envolver muitos aspectos em um ambiente organizacional: sistemas de suporte à decisão, integração de bancos de dados heterogêneos, *warehouses*¹⁷, bibliotecas digitais, sistemas de recuperação da informação, sistemas de correio eletrônico, dentre outros. Tomadores de decisão precisam de informações de fontes distintas, mas em geral não conseguem obter e combinar estes dados em tempo razoável, devido a dificuldades de acessar e integrar sistemas.

Um dos objetivos da inter-operabilidade no ambiente informacional é buscar o aumento do valor da informação no momento em que dados de várias fontes são acessados, relacionados e combinados. Uma simples integração entre um número cada vez maior de fontes disponíveis *on-line* conduz à “sobrecarga de informações” (WIEDERHOLD, 1999). O usuário não consegue mais extrair informação de um conjunto de dados recuperados.

Stuckenschmidt et al. (2000) descrevem a integração como o conjunto dos seguintes aspectos:

- **Integração técnica:** a *Web* proporciona a infra-estrutura para o intercâmbio de grandes quantidades de dados, a partir do acesso a páginas e bancos de dados;
- **Integração sintática:** existem muitos padrões para integração de fontes de dados, desde as clássicas interfaces de bancos de dados, (ODBC-*Object Database Component*), até os padrões desenvolvidos para a *Web* (XML);
- **Integração estrutural:** o primeiro problema que ocorre além do nível sintático é a integração de estruturas heterogêneas, o qual é tratado por sistemas mediadores que definem regras para mapeamento entre diferentes estruturas;
- **Integração semântica:** diz respeito à solução de conflitos semânticos que tornam impossível o mapeamento de conceitos entre duas ou mais fontes.

¹⁷ “Armazém de dados”; trata-se de um grande banco de dados que não utiliza dos dados operacionais e auxilia em processos de apoio à decisão.

Pode-se assim, buscar a inter-operabilidade nos diversos níveis de um ambiente informacional. Em última instância, a inter-operabilidade corresponde às ações que agregam valor aos processos executados por pessoas e/ou sistemas nas quais se abordam questões relativas à combinação de diferentes fontes de dados.

Uma proposta abrangente que pretende partir do nível técnico e atingir níveis mais abstratos, classificando diversas possibilidades de inter-operabilidade é apresentada na Figura 7:

Inter-operabilidade técnica: <i>A maioria dos aspetos relacionados a manter a inter-operabilidade, considera questões técnicas que incluem garantir o desenvolvimento de padrões de comunicação, transporte, armazenamento e representação, tal como o Z.39-50¹⁸, ISO-ILL¹⁹, XML, etc. É necessário que se assegure que estes padrões se encaminhem para uma convergência benéfica, possibilitando que os sistemas possam fazer uso de mais de uma das abordagens de padrões.</i>
Inter-operabilidade Semântica: <i>Apresentam-se questões relativas a como as informações disponíveis de “gateways”²⁰ podem se integrar. Inevitavelmente, essas informações são descritas por termos que representam conceitos similares, ou mesmo termos idênticos significando diferentes coisas, introduzindo confusão e erros..</i>
Inter-operabilidade política / humana: <i>Além das questões sobre a maneira como a informação é descrita ou disseminada, a decisão de fazer recursos internacionalmente distribuídos tem implicações para as organizações (que podem ver isso com uma perda de sua propriedade), seus funcionários (que podem não possuir as habilidades para lidar com sistemas mais complexos) e para usuários finais. O processo de mudança, os funcionários e o seu treinamento não são considerados quando se vai substituir um dado recurso, mas são cruciais para assegurar um uso efetivo e de longo prazo de qualquer serviço.</i>
Inter-operabilidade intercomunidades: <i>Os limites tradicionais entre instituições têm se tornado difusos. Pesquisadores necessitam, cada vez mais, de acesso a informação presente em varias fontes, dentro ou fora de suas próprias áreas de atuação. Complementando os trabalhos no setor de bibliotecas, iniciativas estão ocorrendo para desenvolver comunidades que forneçam informações relacionadas, como por exemplo na área de museus e arquivologia. Os problemas são similares e existem benefícios na adoção de soluções comuns.</i>
Inter-operabilidade internacional: <i>Cada uma das questões apresentadas acima, torna-se mais ampla quando considerada em escala internacional, em que diferentes abordagens técnicas, praticas de trabalho e organização, têm sido desenvolvidas ao longo de anos. Importantes também nesse contexto são as diferenças culturais.</i>

Figura 7 – Classificação para a inter-operabilidade - UK Interoperability Focus

Todas as camadas são importantes (e inclusive remetem às Figuras 1 e 3), mas o tópico de maior interesse no contexto deste trabalho é a inter-operabilidade semântica. De forma mais simples, o que se busca é reduzir diferenças de significado, construindo um painel

¹⁸ ANSI/NISO Z39.50 é um protocolo de comunicação para recuperação da informação em ambientes cliente-servidor, muito utilizado em aplicações de bibliografia e bibliotecas digitais.

¹⁹ ISO 10160:1997 *Interlibrary Loan Application* (ILL), padrões projetados para permitir, com uniformidade técnica mínima com outros padrões, a interconexão de sistemas de diferentes fabricantes, diferentes níveis de complexidade e diferentes tipos de gerenciamento.

²⁰ Dispositivo de *hardware* ou *software* que permite o fluxo de dados entre redes diferentes.

significativo das relações entre as fontes de dados. As ferramentas aqui propostas para solucionar tais problemas são as ontologias, discutidas em mais detalhes no Capítulo 5, onde se propõe um modelo e um meta-modelo. Compreender um pouco melhor o assunto “ontologias” é o objetivo do Capítulo 4.

2.4 Que nível de integração é necessário?

Em geral, os problemas relativos à integração de fontes heterogêneas dizem respeito à grande diversidade entre fontes de informação disponíveis, à semântica distinta entre aplicações, a dados armazenados em sistemas legados, às diversas granulometrias²¹ possíveis de se extrair de uma fonte de informação, ao tamanho das aplicações e à evolução constante das fontes de informação e sistemas. Cada um desses aspectos é tratado separadamente em seguida.

De forma mais específica, a heterogeneidade entre fontes de informação tem origem, principalmente, na sua diversidade e em questões semânticas. As informações estão armazenadas em diferentes formatos, esquemas, vocabulários e ontologias. Também a maneira pela qual a informação é acessada pode variar em função de interfaces, linguagens de consulta, protocolos, etc.

No caso de diferenças semânticas entre as aplicações, deve-se observar que os programas que criam e mantêm dados são a chave para a compreensão do problema. Os sistemas têm regras implícitas complexas em relação à interpretação e manutenção dos dados, as quais, em geral, estão codificadas e são de difícil compreensão. Além disso, estes sistemas são de naturezas variáveis, incluindo sistemas baseados em conhecimento, sistemas comerciais, bibliotecas digitais, repositórios de softwares, dentre outros.

O acesso a fontes de dados legados apresenta problemas similares aos problemas semânticos. Nas grandes aplicações médicas, de seguros, bancárias, de telecomunicações, etc, o acesso a bancos de dados legados exige que se compreenda e se reutilize partes de aplicativos que definem o acesso aos dados. Pode ser difícil deduzir a semântica inserida em um sistema legado e quase impossível fornecer interpretações a partir de um exame externo.

Nem sempre se faz uso de uma fonte de informação completa. Esta é a origem dos problemas relativos às granulometrias de informação possíveis ao se acessar uma fonte. Componentes internos (por exemplo partes de bancos de dados e partes de grandes sistemas) devem ser extraídos e integrados para se obter a informação necessária.

²¹ Refere-se a tamanho básico das unidades que podem ser manipuladas; habilidade de conceitualizar o mundo em diferentes níveis de complexidade.

Finalmente, o tamanho dos sistemas atuais e a constante evolução das fontes de informação adicionam mais complexidade ao processo: torna-se necessário que muitas fontes sejam consultadas, interpretadas e coordenadas ao se propor um sistema integrado. Além disso, é obrigatório que isso seja feito de forma dinâmica, em função da constante evolução dos sistemas e das fontes de dados que os alimentam.

Em abordagens de cunho mais prático, alguns autores citam situações comuns, a partir das quais pode-se entender que nível de integração se pretende obter ao combinar fontes heterogêneas. Paepcke (1996) apresenta um cenário, em uma empresa industrial, onde um engenheiro tem dificuldades para reunir material sobre uma determinada peça, fabricada há algum tempo pela empresa. Fica claro que a dificuldade em recuperar informações está relacionada também, além das tradicionais diferenças nos formatos de representação e armazenamento, à produção de dados em diferentes períodos de tempo e para diferentes finalidades.

Outro exemplo prático é citado por Gupta et al. (2000), que apresentam uma situação em um domínio específico. A necessidade de integração de informação é essencial em biologia, em função da grande heterogeneidade nos tipos de dados envolvidos, e, em função da diversidade de abordagens (fisiológica, anatômica, bioquímica, etc.) de que os biólogos necessitam para estudar um mesmo fenômeno. É importante integrar estes dados, os quais estão armazenados em diferentes fontes, cada qual com características próprias.

Ainda como exemplo, Allen et al. (2002) citam problemas encontrados na área de integração de imagens de coleções de museus, que dificultam o intercâmbio de dados e a execução de consultas inter-coleções.

Pode-se também lembrar de exemplos práticos como os dicionários de dados, construídos pela análise de sistemas, e as linguagens artificiais e listas padronizadas, na biblioteconomia. São abordagens eficientes para ambientes restritos, mas não são adequadas a ambientes abertos e heterogêneos, característicos dos grandes sistemas distribuídos.

O objetivo de um projeto de integração, ou seja, os resultados que se pretende obter, os propósitos e as condições externas que limitam esse projeto, são os maiores determinantes do nível de integração necessário. Mas, pode-se afirmar que o problema de integrar fontes diversas se apresenta em várias dimensões, não estando restrita apenas à recuperação da informação. Para se conseguir uma integração efetiva devem-se executar tarefas relacionadas à recuperação, interpretação, gerenciamento local dos dados obtidos e compartilhamento (PAEPCKE, 1996). Só assim um projeto de integração poderá agregar valor ao trabalho das pessoas, integrando fontes internas e externas, de diferentes formatos, com diferentes formas de acesso e em dimensões

distintas de tempo e espaço.

O nível de integração desejado será assim determinado, em última instância, pelas necessidades dos usuários. Observa-se, entretanto, que estas necessidades tem mudado com o surgimento de novos ambientes que tornam disponíveis novas fontes de dados, de maior complexidade. Novas técnicas devem ser criadas de forma a abranger o novo ambiente de dados.

Ao se procurar entender que nível de integração é procurado e quais são as possibilidades para se estudar a inter-operabilidade, pode-se também utilizar algum tipo de classificação, com a finalidade de reduzir a complexidade do problema e facilitar sua compreensão. Algumas propostas são citadas na literatura para tal classificação e serão apresentadas na seção seguinte, com o objetivo de possibilitar melhor entendimento dos componentes que são comuns em um projeto de integração. Com esta classificação será possível confrontar duas ou mais soluções de inter-operabilidade, e, talvez, através de métricas, escolher a mais viável.

2.5 Classificação de elementos em modelos de inter-operabilidade

Com o objetivo de minimizar a complexidade do problema, alguns autores propõem estruturas de classificação que tentam facilitar a compreensão das variáveis envolvidas. Além disso, a existência de uma classificação pode proporcionar uniformidade às diversas abordagens. Apresentam-se em seguida algumas tentativas de realizar tal classificação obtidas na literatura.

Problemas de inter-operabilidade rapidamente detectados podem se tornar parte do conjunto de decisões ainda na etapa de projeto, resultando em uma implementação de sistemas com menos problemas. Os componentes principais de um projeto de integração são (Keshav et al., 1999):

- O “*translator*”, que converte dados e funções entre os diferentes formatos dos componentes, sem alterar o conteúdo da informação;
- O “*controller*”, que coordena e intermedia o movimento da informação entre os componentes usando processos pré-definidos de tomada de decisão;
- O “*extender*”, que adiciona novas características e funcionalidades, proporcionando melhorias às capacidades de um componente;
- Os elementos híbridos: “*translator-controller*”, o “*translator-controller-extender*” e o “*controller-extender*”.

As categorias das ferramentas que tornam viáveis projetos de inter-operabilidade

podem ser classificados como “*enablers*” (OUKSEL; SHETH, 1999). Os principais “*enablers*” e suas funções são:

- Transparência de terminologia: permite ao usuário escolher as ontologias de seu interesse e combiná-las em diferentes domínios do conhecimento;
- Processamento sensível ao contexto: o sistema de informação deve reconhecer o contexto e usá-lo para limitar a sobrecarga de informação sobre o usuário;
- Regras de interação: trata-se de uma especificação formal, construída através da semântica, sobre o formato das mensagens e dos tipos de dados, sem interferência no conteúdo que está sendo comunicado;
- Correlação semântica: permite a representação de informações semanticamente relacionadas, sem levar em conta sua distribuição, proporcionando o uso de informações relevantes em qualquer local ou período de tempo.

ARPA Intelligent Integration of Information Reference Architecture (1995) também apresenta um conjunto de serviços cujo objetivo é classificar ferramentas utilizadas por produtores de informação para implementar soluções para a inter-operabilidade em um domínio específico:

- Serviços de processamento de consultas;
- Serviços de integração (incluem serviços de tradução de linguagens, reestruturação de dados e serviços de mediação especializados);
- Serviços de intercâmbio de dados e conhecimento;
- Serviços de privacidade e segurança;
- Serviços de gerenciamento de recursos;
- Serviços de repositórios.

Após classificar e compreender o sistema, pode-se tentar concluir se o nível de integração desejado foi alcançado. Uma vez identificados os problemas e classificados os principais componentes do sistema, é importante também mensurar a eficácia e a eficiência das soluções adotadas. Na literatura também podem ser encontradas propostas para métricas utilizadas em projetos de inter-operabilidade.

A definição de métricas é importante na medida de rendimento de qualquer sistema e na verificação do nível de satisfação em relação às respostas obtidas. Tentativas de estabelecer métricas para a inter-operabilidade são bastante genéricas e talvez não possibilitem a obtenção de resultados práticos (SELIGMAN, 1995).

Para se que possa ter uma idéia da variedade de possibilidades, de teorias e projetos que buscam a integração de fontes de dados, algumas propostas de diferentes comunidades científicas são apresentadas na seção seguinte.

2.6 Pesquisas sobre inter-operabilidade

As abordagens para se buscar inter-operabilidade são variadas e podem também ser consideradas técnicas combinadas. Existem pesquisas sobre a inter-operabilidade, provenientes de diferentes comunidades acadêmicas. Encontram-se projetos realizados por pesquisadores da área de bancos de dados, de bibliotecas digitais, de recuperação da informação, dentre outros. Em cada uma dessas áreas têm-se abordagens diversificadas para a solução do problema. Também se encontra material relacionado em projetos de desenvolvimento de sistemas públicos integrados.

Existem ainda abordagens que trabalham estritamente dentro dos limites da tecnologia, como é o caso do que se convencionou chamar de *middleware*. Com a popularidade dos sistemas abertos, da Internet e aplicativos que trabalham fora dos limites da organização, tornou-se relevante o estudo de uma camada de software que execute a integração dos componentes distintos e oculte a heterogeneidade inerente aos sistemas. Essa camada é o *middleware*, que faz a intermediação entre os aplicativos e o sistema operacional. A inter-operabilidade nas empresas, que consiste na integração de processos através de sistemas de informação é estudada por uma nova disciplina, denominada *Enterprise Application Integration (EAI)*.

A abordagem da comunidade de inteligência artificial trata do assunto utilizando-se de três ferramentas básicas para representação do conhecimento: estruturas para a organização de bases de dados e conhecimento (ontologias); a lógica (como forma de representação do conhecimento que pode ser implementada em computadores) e a computação (para implementação de sistemas condizentes com as necessidades atuais). Essa abordagem possui relação estreita com a solução adotada nesse trabalho.

Outra comunidade que produz trabalhos relevantes sobre integração é a que pesquisa bibliotecas digitais. As bibliotecas digitais agregam recursos provenientes de várias fontes, e portanto, correspondem a um ambiente em que a busca pela inter-operabilidade é constante. A literatura na área relativa a problemas de integração é extensa.

Ontologias são um tópico relevante para estudos de inter-operabilidade, visto que, determinar conceitos e relações em domínios específicos de conhecimento é parte importante de um projeto de integração entre fontes heterogêneas. O desenvolvimento de ontologias, bem como

a possibilidade de combiná-las, para alcançar níveis de integração desejados, é atualmente uma importante atividade de pesquisa interdisciplinar, que integra a Ciência da Informação e a Ciência da Computação. Ambientes de informação distribuídos em redes são cada vez mais dominantes na maioria das atividades relacionadas à informação para negócios, para serviço público e para uso pessoal. A integração semântica nestes contextos pode ser conseguida através de ferramentas que utilizam ontologias para organizar o conhecimento.

No Capítulo 3 é apresentada uma introdução à representação do conhecimento, em que se pesquisa como computadores devem ser capazes de interpretar significado. Essa possibilidade abre o caminho para que possam inter-operar (sintática e semanticamente) sistemas computacionais representados por diferentes modelos e associados a fontes de dados distintas.

2.7 Considerações finais

Algumas vezes as abordagens apresentadas são aplicadas em conjunto, como parte de projetos mais abrangentes. O Capítulo 5 apresenta a construção de um meta-modelo, passível de aplicação em diversos ambientes, que busca reduzir conflitos semânticos e possibilitar a inter-operabilidade de fontes heterogêneas. Para lidar com essas diferenças semânticas, o modelo prioriza a abordagem de ontologias, eliminando limitações semânticas encontradas nos ambientes informacionais digitais. O ambiente em que esse modelo será testado é o projeto SABiO.

Antes de estudar o modelo para a inter-operabilidade, deve-se abordar com maiores detalhes o que são ontologias, como são desenvolvidas e usadas para a integração. O Capítulo 3 introduz o assunto, tornando mais claro o tipo de modelo que será proposto e como funcionam seus componentes. As ontologias e sua utilidade para inter-operabilidade são apresentadas em mais detalhes no Capítulo 4, em que também são citados exemplos de trabalhos relacionados.

CAPÍTULO 3 – ONTOLOGIAS

Para entender como o uso de ontologias pode auxiliar na tarefa de prover interoperabilidade, alguns conceitos básicos devem ser apresentados. Essa seção introduz tais conceitos, mostra os diferentes significados da palavra “ontologia” e enumera seus componentes típicos. São apresentados ainda os tipos de ontologias mais comuns, onde elas podem ser encontradas e uma breve pesquisa (metodologias, ferramentas, linguagens e exemplos) sobre como construí-las. Não se tem a pretensão de cobrir toda a literatura sobre ontologias e nem todas as possibilidades existentes para a integração via ontologias. O que se espera apresentar é uma visão geral do assunto, importante no escopo deste trabalho.

3.1 Sobre representação do conhecimento

Os problemas relativos à representação da informação e do conhecimento são abordados por estudiosos de diversas áreas (CAMPOS, 1995). No contexto da Ciência da Informação os problemas em representação de conteúdo ocorrem à medida que o volume de documentos a processar e a ordenar cresce, de forma que os usuários não mais se contentam com a organização por grandes classes e exigem informações mais precisas (ROBREDO, 1986). Com o grande volume de material disponível atualmente, o problema se torna crítico. A dificuldade em localizar e recuperar uma grande quantidade de documentos levou ao desenvolvimento de novas técnicas de análise e representação documentária.

Neodobity²² (1986) citado por Campos (1995) identifica princípios da Teoria Geral da Terminologia que podem ser utilizados em *sistemas especialistas*²³ e em inteligência artificial, áreas que lidam com conceitos, sistemas de conceitos e relações entre conceitos. Estudos nessas áreas têm sido desenvolvidos visando à elaboração de sistemas baseados em conceitos. Um exemplo da utilização de uma estrutura conceitual para executar tarefas em um domínio específico do conhecimento são as ontologias, as quais são estudadas nas seções subseqüentes.

Capturar o conhecimento humano e torná-lo explícito não é tarefa simples. O problema toma proporções ainda maiores quando se pretende registrar a experiência humana e representá-la sobre a forma de *softwares* (MENDES, 1997). Em se tratando de sistemas digitais utilizados em computadores, é necessário ainda codificá-lo em algoritmos e estruturas de dados.

²² NEODOBITY, W. Classification, artificial intelligence and cognitive psychology In: GAUL, W.; SCHADER, M. *Classification as a tool for research*. Amsterdam: Elsevier, 1986. p. 245-352.

²³ Sistemas que procuram substituir a *expertise* de especialistas humanos em determinadas situações, estudados na computação, pela Inteligência Artificial.

Para tornar o conhecimento implícito acessível ao computador, são construídos *sistemas baseados em conhecimento*²⁴ com a utilização de *linguagens declarativas*²⁵, cuja forma de expressão é mais próxima do idioma humano. Tais sistemas ajudam a expressar o conhecimento em uma forma inteligível, tanto pessoas quanto para computadores.

Nessa abordagem, a representação do conhecimento se desenvolveu como um ramo da inteligência artificial, associado ao desenvolvimento de *sistemas especialistas*. Atualmente, muitas outras atividades, além da inteligência artificial, exigem a utilização de conhecimento humano codificado, como por exemplo: a recuperação da informação, o projeto de circuitos, realidade virtual, reconhecimento de voz, dentre outros. Assim, analisar o conhecimento e mapeá-lo para a forma digital é hoje uma tarefa de importância para diversas áreas do conhecimento.

A representação do conhecimento em sistemas digitais diz respeito à aplicação de lógica no desenvolvimento de modelos computacionais para um determinado domínio do conhecimento (SOWA, 1999b). Essa idéia é compartilhada por Nebel (2001) que afirma que o problema de representar o conhecimento pode ser abordado pelo uso de uma grande variedade de lógicas.

A representação do conhecimento é uma matéria multidisciplinar que aplica teorias e técnicas da lógica, ontologias e computação. A lógica fornece a estrutura formal e *regras de inferência*²⁶; as ontologias definem “os tipos das coisas” e as “coisas” que existem no domínio da aplicação, para que possam ser representadas; a computação trata das aplicações e construção de sistemas (SOWA, 1999a).

Sem a lógica, a representação do conhecimento não tem critérios para determinar se declarações são contraditórias ou redundantes. Sem as ontologias, não se podem definir precisamente os termos e símbolos ou compartilhar seu significado. Sem os modelos computacionais, a lógica e a ontologia não podem ser implementadas em um computador.

Um aspecto fundamental em sistemas computacionais, que atuam em *bases de conhecimento*, é a possibilidade de se executarem *inferências*. Através de uma “*máquina de inferência*”²⁷, os fatos, regras e heurísticas que compõem a *base do conhecimento*, são aplicadas à

²⁴ Sistemas que consultam uma base de conhecimento.

²⁵ Ao contrário de uma linguagem procedural (que especifica uma seqüência de passos para produzir um resultado) uma linguagem declarativa descreve relacionamentos entre variáveis em termos de funções ou regras de inferência.

²⁶ Inferência é o processo lógico no qual novos fatos são derivados de fatos já conhecidos, através da aplicação de regras.

²⁷ Um programa que infere novos fatos a partir de fatos conhecidos, utilizando-se de regras. Em geral faz parte de um interpretador *Prolog*, sistema especialista ou sistema baseado em conhecimento.

solução de um problema (MENDES, 1997). Os serviços de *inferência* são abordados brevemente no Anexo B.

Todas as principais formas de representação do conhecimento em sistemas computacionais estão relacionadas à lógica: regras, *redes semânticas*, *linguagens orientadas a objeto*, *Prolog*, Java²⁸, *SQL-Structured Query Language*, *Redes de Petri*²⁹ e *KIF-Knowledge Interchange Format*³⁰. Duas notações lógicas básicas, equivalentes em sua semântica, usadas na ciência da computação são o cálculo de predicados³¹ e os grafos conceituais³².

As diferenças entre as ontologias aparecem por variações culturais, geográficas, técnicas, dentre outras. Segundo Sowa (1999b), o problema de “alinhar” os termos de ontologias diferentes é essencialmente o mesmo que se tem ao comparar palavras de vocabulários de linguagens naturais diferentes.

Através das ontologias, no contexto desse trabalho, procura-se tornar explícito o conhecimento presente em fontes bibliográficas de dados, para que ele possa ser manipulado por sistemas e apresentado de forma útil e integrada ao usuário.

Para que se possa entender o significado geral para o termo “ontologia” e como ele é utilizado no contexto em estudo, a seção seguinte apresenta uma introdução ao assunto, com definições, características, tipos e exemplos.

3.2 Sobre as Ontologias

3.2.1 Relação entre ontologias, lógica e organização do conhecimento

O termo ontologias tem um sentido especial em computação e inteligência artificial, diferente daquele tradicional, adotado na filosofia³³.

De forma geral, estudar ontologias é definir categorias para as coisas que existem em um mesmo domínio. 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 determinada linguagem (SOWA, 1999b).

²⁸ Linguagem de programação orientada a objeto, de propósito geral, desenvolvida pela *Sun Microsystems*.

²⁹ Grafo direto bipartido onde os nós são locais (representados por círculos) ou transições (representadas por retângulos). São usados por exemplo, para modelar protocolos de rede.

³⁰ Linguagem de programação orientada a objeto que objetiva o intercâmbio de conhecimento entre programas distintos.

³¹ O cálculo de predicados é uma notação tradicional para a lógica.

³² Grafos conceituais são uma forma bidimensional de lógica que baseada em redes semânticas.

³³ Maiores detalhes na Seção 3.2.2.

Historicamente o termo ontologia tem origem no grego “*ontos*”, ser e “*logos*”, palavra. É um termo novo na história da filosofia, introduzido originalmente com o objetivo de distinguir o estudo do ser como tal (ou seja do ser humano em sua essência), do estudo dos vários tipos de outros seres das ciências naturais. O termo original é a palavra aristotélica “*categoria*”, que pode ser usada para classificar e caracterizar alguma coisa.

Em seus tratados, Aristóteles apresenta categorias que servem de base para classificar qualquer entidade e introduz ainda 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 e é fundamental para estudos semânticos, desde dicionários até a *Web Semântica*³⁴ (SOWA, 1999b).

Outra importante contribuição aristotélica foi a invenção dos *silogismos*³⁵ como um padrão formal para representar *regras de inferência*. Esses estudos são a base para as regras de herança em hierarquias e para deduções sobre restrições em uma hierarquia. Também são a base da *lógica de primeira ordem*³⁶ que é a base da *lógica descritiva*, da qual duas versões modernas são o *DARPA Agent Markup Language* (DAML)³⁷ e o OIL³⁸ (SOWA, 1999b). A *lógica descritiva* e o OIL são utilizados no Capítulo 5 para a representação do meta-modelo e do processo de integração³⁹. A referência a esses assuntos neste trabalho limita-se aos aspectos relevantes para desenvolvimento do meta-modelo. São temas extensos e aprofundar-se neles foge aos objetivos aqui propostos.

Recentemente, a necessidade de padronizar a codificação do conhecimento presente nos sistemas levou a *American National Standards Institute* (ANSI) a elaborar uma primeira proposta em que todo o conhecimento sobre um domínio fosse reunido em um esquema conceitual (ver Figura 8). A interface com o usuário possibilita requisições ao banco de dados onde estão aplicações que armazenam e compartilham dados. O esquema conceitual (representando o conhecimento comum) liga os três círculos fornecendo definições comuns das entidades da aplicação e dos relacionamentos entre elas. A Figura 8 mostra as inter-relações em

³⁴ Trata-se de uma extensão da *Web* atual na qual a informação é apresentada com um sentido bem definido, possibilitando que pessoas e computadores trabalhem juntos³⁴.

³⁵ Raciocínio dedutivo; argumento ou raciocínio formado de três proposições - a maior, a menor (premissas) e a conclusão, sendo esta última deduzida da maior, por intermédio da menor.

³⁶ Uma linguagem que descreve a verdade dos fatos em formulas matemáticas. As formulas descrevem propriedades de termos e tem um valor verdadeiro.

³⁷ A linguagem DAML tem sido desenvolvida com uma extensão do XML e RDF e seu objetivo é proporcionar ferramentas para facilitar a concepção da *Web Semântica*.

³⁸ Para maiores detalhes sobre essa linguagem ver Anexo C.

³⁹ Mais informações e terminologia sobre a lógica descritiva são apresentados no Anexo B.

um sistema de integração:

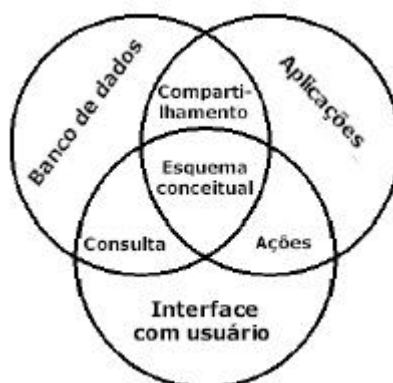


Figura 8 – Esquema conceitual como núcleo de um sistema integrado de dados (SOWA, 1999b)

O esquema conceitual é importante para o projeto, desenvolvimento e uso de aplicações integradas e é a base de várias outras funcionalidades (*linguagens de quarta geração*⁴⁰, *sistemas orientados a objeto*⁴¹ e as *ferramentas CASE*⁴²). A tentativa mais recente de integrar conhecimento é a *Web Semântica*. Sua maior contribuição foi propor o XML como uma sintaxe comum. Entretanto, o problema está em desenvolver uma semântica compatível entre fontes e sistemas heterogêneos⁴³. Significado é facilmente interpretado por pessoas, mas computadores têm grandes limitações nessa tarefa.

Atualmente, pesquisadores e profissionais envolvidos com sistemas entendem que para obter melhorias de desempenho, devem ser também tratados aspectos relativos à obtenção de dados e como estes são apresentados às pessoas. Segundo Paepcke (1996), apenas a busca não é mais suficiente e outras tarefas também devem ser executadas simultaneamente, durante o processo de fornecer informação útil para o usuário. Dessa forma, muita importância tem sido dada à organização das bases de dados, procurando melhorar o processo de recuperar a informação e apresentá-la. Nesse contexto, atenção tem sido dispensada a problemas que dificultam a inter-operabilidade.

Para utilizar uma base de dados, o conhecimento ali contido deve se ser explicitado, e em seguida, compreendido. Expressar qualquer coisa em uma representação requer uma

⁴⁰ Linguagens de alto nível não procedurais, construída junto a sistemas de bancos de dados.

⁴¹ Sistemas desenvolvidos com o uso de linguagens de programação que utilizam o conceito de “objeto”, o qual é uma estrutura de dados (tipo abstrato) encapsulado junto a rotinas chamadas métodos, os quais operam sobre os dados.

⁴² *Computer-aided software engineering*, são softwares cujo objetivo é auxiliar o projeto de outros softwares.

⁴³ Esse assunto também é abordado nas Figuras 5 e 6 e explicações na Seção 2.2.

linguagem única, que pode ser uma linguagem natural⁴⁴ ou lógica, a qual expresse a linguagem natural em termos computacionais (SOWA, 1999b). O problema de representar sistemas diferentes em lógica é complexo e, desde que a semântica da lógica de primeira ordem foi estabelecida⁴⁵, a única coisa padronizada até hoje foi a notação. Propondo alternativas, o *Comitê de Metadados NCITSL8*⁴⁶ desenvolveu duas notações diferentes, utilizando como base uma semântica única:

- **KIF - *knowledge interchange format***: é uma notação para lógica, de sintaxe simples e um grupo de caracteres restrito, que pretende intermediar dois sistemas de computação heterogêneos;
- **CGs - *Conceptual Graphs***: é uma notação para lógica, baseada em *grafos*, que utiliza também aspectos lingüísticos e de redes semânticas; trata-se de uma linguagem que apresenta a lógica de forma legível para as pessoas.

Para as aplicações relacionadas à computação, a atenção está voltada para a obtenção de axiomas⁴⁷ precisos, utilizados no processamento. A Ciência da Informação enfatiza as pessoas, as quais têm a responsabilidade central de processar o conhecimento. Assim, a preocupação nessa área está voltada para a padronização da terminologia usada para encontrar e classificar a informação. Para tal, as disciplinas relativas a Tratamento da Informação, que herdaram técnicas da biblioteconomia, são relevantes na busca por inter-operabilidade. Mesmo em ambientes digitais, normalmente estudados por pesquisadores da computação, essas técnicas são essenciais para se integrarem diferentes fontes de dados.

3.2.2 Definições para ontologia

Existem diferentes sentidos para o termo e uma breve discussão sobre o seu significado pode esclarecer sua utilização no contexto desse trabalho. Além disso, é importante diferenciar seu uso no campo da filosofia e no campo dos sistemas computacionais.

Na filosofia, uma das definições obtida no dicionário *Oxford* de Filosofia para ontologia é “[...] o termo derivado da palavra grega que significa ‘ser’, mas usado desde o século

⁴⁴ Uma linguagem falada ou escrita por humanos. A compreensão da linguagem natural é um dos maiores problemas para a interpretação por computadores, em função de sua irregularidade, complexidade e problemas de significado.

⁴⁵ Alfred Tarki, em 1935, produziu uma teoria de referência (semântica) à qual relaciona objetos na lógica a objetos do mundo real.

⁴⁶ Comitê técnico do NCITS (*National Committee on Information Technology Standards*) que objetiva estabelecer padrões para especificar e descrever dados, buscando mais facilidade para o uso humano e para permitir processamento inteligente por computadores.

⁴⁷ Princípio evidente, que não precisa ser demonstrado; máxima, sentença; norma admitida como princípio.

XVII para denominar o ramo da metafísica que diz respeito àquilo que existe”⁴⁸. O Dicionário Aurélio traz as seguintes definições: “1 Ciência do ser em geral. 2 Parte da metafísica que estuda o ser em geral e suas propriedades transcendentais”⁴⁹.

Merriam-Webster e Gove (2002)⁵⁰ citado por Corazzon (2002, p. 1) apresentam um sentido para o termo que diz respeito a sua utilização pela computação, mais especificamente pela comunidade de inteligência artificial: “uma teoria que diz respeito a tipos de entidades e, especificamente, a tipos de entidades abstratas que são aceitas em um sistema com uma linguagem”. Nesse sentido, o termo é usado na pesquisa de representação do conhecimento em inteligência artificial.

Mesmo considerando-se apenas o domínio da computação, são diversos os conceitos apresentados para o termo e existem muitas contradições. Uma das definições mais conhecidas para ontologias no campo da Inteligência Artificial é a de Gruber (1996)⁵¹ citado por Corazzon (2002, p. 1):

Uma ontologia é uma especificação explícita de uma conceitualização⁵². O termo é tomado da filosofia, onde uma ontologia é uma consideração sistemática da existência. Para os sistemas da inteligência artificial, o que existe é o que pode ser representado. [...] Dessa forma, no contexto da inteligência artificial, pode-se descrever a ontologia de um programa definindo um grupo de termos representacionais. Em tal ontologia, definições associam nomes de entidades no universo do discurso (por ex. classes, relações, funções ou outros objetos, com textos legíveis para pessoas, descrevendo o que os nomes significam, e, os axiomas formais que restringem a interpretação e o uso adequado desses termos) [...].

Esta definição é questionada por Guarino e Giaretta (1995)⁵³ citado por Corazzon (2002, p. 1), que, após examinar algumas possíveis interpretações para o termo ontologia, afirmam:

[...] um ponto inicial nesse esforço de tornar claro o termo, será uma análise cuidadosa da interpretação adotada por Gruber. O principal problema com tal interpretação é que ela é baseada na noção conceitualização, a qual não corresponde a nossa intuição. [...] Uma

⁴⁸ BLACKBURN; S. Consultoria da edição brasileira. MARCONDES, D. Tradução MURCHO, D. et al. *Dicionário Oxford de Filosofia*. Rio de Janeiro; Jorge Zahar, 1997.

⁴⁹ FERREIRA, A. B. H. *Novo Aurélio: o dicionário da língua portuguesa*. 3. ed. Rio de Janeiro: Nova Fronteira, 1999.

⁵⁰ MERRIAM-WEBSTER; GOVE, P. B. *Webster's Third New International Dictionary*. Unabridged. New York: Merriam-Webster, 2002. 2.783 p.

GRUBER, T. (1996). *What is an Ontology?* Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 14 set. 2002.

⁵² Conceito introduzido por GENESERETH, Michael R.; NILSSON, L. *Logical foundation of AI*. San Francisco, Los Altos, Califórnia: Morgan Kaufman, 1987; assim que esta no paper a coleção de objetos, conceitos e outras entidades que se assume existir em alguma área de interesse e os relacionamentos entre elas. Uma conceitualização é uma visão abstrata e simplificada do mundo que se deseja representar. Escolher a conceitualização é o primeiro passo para a representação do conhecimento.

⁵³ GUARINO, N.; GIARETTA, P. (1995). *Ontologies and KBs, towards a terminological clarification*. Disponível em: <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/KBKS95.pdf>>. Acesso em: 13 jul. 2002.

conceitualização é um grupo de relações extensionais descrevendo um ‘estado das coisas’ particular, enquanto a noção que temos em mente é uma relação intensional, nomeando algo como uma rede conceitual a qual se superpõe a vários possíveis ‘estados das coisas’⁵⁴.

Guarino (1998) propõe uma revisão da definição de conceitualização que faz uso deste *aspecto intensional*, para que se possa obter uma interpretação satisfatória da definição de Gruber (1996, p. 1):

[...] por outro lado, em sua utilização em inteligência artificial, ontologia se refere a um artefato de engenharia, constituído por um vocabulário específico usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário. Este conjunto de fatos tem a forma da teoria da lógica de primeira ordem, onde as palavras do vocabulário aparecem como predicados unários ou binários. [...] Duas ontologias podem ser diferentes no vocabulário usado (palavras em inglês, francês ou italiano) e mesmo assim compartilhar a mesma conceitualização.

Borst (1997, p. 12) apresenta uma definição considerada mais simples e completa: “Uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Nessa definição, “formal” significa legível para computadores; “especificação explícita” diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas que são explicitamente definidos; “compartilhado” quer dizer conhecimento consensual; e, “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real.

Além disso, a ontologia também deve definir as regras que regulam a combinação entre os termos e as relações. As ontologias e os relacionamentos entre os termos são criados por especialistas naquele domínio. Os usuários formulam consultas usando os conceitos especificados na ontologia e o sistema trata da heterogeneidade e da distribuição pelos repositórios. Em última instância, uma ontologia define uma “linguagem” (conjunto de termos) que será utilizada para formular consultas.

Discussões mais aprofundadas sobre o assunto podem ser encontradas em Guarino e Giaretta (1995). Outras definições e considerações para o termo são encontradas em Albertazzi (1996), Neches et al. (1991), Wache et al. (2001), Uschold e Gruninger (1996) e Chandrasekaran, Johnson, Benjamins (1999). Para uma discussão mais detalhada sobre o termo, considerações sobre seu sentido e críticas a diversos conceitos, uma boa referência é Guarino (1996). Para entender melhor as definições de ontologia em filosofia e Inteligência Artificial, uma referência é Guarino (1998). Em outra abordagem, Ozkural (2001) refere-se a ontologias como uma teoria de classificação. Guarino e Giaretta (1995)⁵⁵ citados por Gomes-Perez, Benjamins (1999),

⁵⁴ Aspectos intensionais e extensionais serão utilizados na integração e são explicados no Anexo A.

⁵⁵ GUARINO, N.; GIARETTA, P. *Ontologies and KBs, towards a terminological clarification*. Disponível em: <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/KBKS95.pdf>>. Acesso em: 13 jul. 2002.

apresentam vários sentidos para o termo e os respectivos níveis de abstração.

Ainda assim, conforme citado, Gruber (1993) elaborou uma definição que se tornou a mais referenciada na literatura da computação. No contexto desse trabalho, não se faz opção por uma definição específica, dentre as diversas apresentadas, nem se pretende discutir aqui qual seria a melhor e mais completa. O termo é utilizado ao longo do trabalho no âmbito de sistemas automatizados.

Conforme já apresentado, a idéia de que técnicas de organização, como o uso de ontologias, podem auxiliar na construção da inter-operabilidade tem sido bastante discutida. A Em seguida, a Seção 3.2.3, apresenta algumas características das ontologias. No Capítulo 4, a Seção 4.3 apresenta aplicações que objetivam a inter-operabilidade e utilizam as ontologias para tal.

3.2.3 Características de uma ontologia

Para que se conheça uma ontologia é preciso ter a capacidade de avaliá-la. Conhecendo-se as características comuns das ontologias, pode-se eliminar ou reduzir obstáculos a seu uso em conjunto. Esta seção apresenta as características das ontologias a partir das citações de diversos autores, de forma que se possa compreendê-las.

Gómez-Perez (1999) apresenta critérios que podem ser utilizados para avaliar ontologias. Os passos apresentados para a avaliação estão relacionados às definições que compõe a ontologia:

- **Verificar a estrutura ou arquitetura da ontologia:** as definições são construídas seguindo os critérios de projeto do ambiente em que são usadas?
- **Verificar a sintaxe das definições:** existem estruturas ou palavras-chave sintaticamente incorretas nas definições?
- **Verificar o conteúdo das definições:** o que a ontologia define ou deixa de definir? O que define incorretamente? O que pode ser inferido e o que não pode?

Observa-se que as ontologias não apresentam sempre a mesma estrutura, mas que algumas características e componentes básicos estão presentes em grande parte delas. Para se identificar uma ontologia e suas características, Gruber (1996) distingue seus componentes básicos:

- **Classes:** nas ontologias, elas são organizadas em uma taxonomia;
- **Relações:** representam o tipo de interação entre os conceitos de um domínio;

- **Axiomas:** são usados para modelar sentenças sempre verdadeiras;
- **Instâncias:** são utilizadas para representar elementos específicos (ou seja, os próprios dados);

Noy e Guinness (2001) também discutem a caracterização dos termos de uma ontologia. Os componentes de ontologias mais conhecidos e suas características são bem similares aos componentes apresentados em Gruber (1996).

Conhecendo os componentes e suas características torna-se possível avaliar a estrutura de uma ontologia, o que é importante para a tarefa de correlacioná-las. Ao se identificar e comparar ontologias para uma tentativa de integração, algumas questões devem ser discutidas e questões básicas podem ser formuladas para entender o contexto Gomez-Pérez, 1999:

- Quais linguagens estão disponíveis e em quais ontologias?
- Quais são os mecanismos para interagir com as ontologias?
 - Qual é o formalismo de representação do conhecimento utilizado?
 - Qual é a infra-estrutura de custos necessária para usar uma ontologia?
- A ontologia é bem documentada?
- A ontologia foi avaliada sobre o ponto de vista técnico?

De forma análoga, mas especificamente no domínio da biologia molecular, Kohler e Schulze-Kremer (2002) listam critérios para que ontologias possam integrar bancos de dados. Alguns critérios são genéricos e podem ser aproveitados para outros domínios. Omelayenko (2001) também apresenta características necessárias para que ontologias possam ser integradas e utilizadas comercialmente em organizações.

Uma ontologia pode assumir várias formas, mas sempre inclui um vocabulário de termos e a especificação de seu significado. O grau de formalidade em que o vocabulário é criado e como se especifica seu significado, classifica as ontologias em categorias (USCHOLD; GRUNINGER, 1996):

- **Altamente informal:** expressa livremente em linguagem natural;
- **Semi-informal:** expressa em linguagem natural de forma restrita e estruturada;
- **Semi-formal:** expressa em uma linguagem artificial definida formalmente;
- **Rigorosamente formal:** termos definidos com semântica formal, teoremas e provas.

Gerar ontologias é uma tarefa trabalhosa e, dessa forma, elas devem ser importadas

sempre que possível (Kohler; Schulze-Kremer, 2002). Outra tarefa trabalhosa, mas necessária, é manter as ontologias atualizadas, importando-as novamente sempre que são atualizadas. Na Seção 3.3 apresentam-se algumas técnicas para se construírem ontologias, nos casos em que não é possível reaproveitar uma já desenvolvida.

Mesmo apresentando propriedades distintas, é possível, conforme citado anteriormente, identificar algumas características comuns a todas as ontologias. Observando grupos de ontologias com mais atenção, pode-se verificar a existência de tipos bem definidos, os quais serão apresentados na seção seguinte.

3.2.4 Tipos de Ontologias

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 cada uma das classes definidas. Esta seção apresenta um conjunto de citações de autores, de onde se podem obter informações sobre os tipos mais comuns de ontologias.

Um estudo abrangente sobre os tipos de ontologias foi desenvolvido por Van-Heijst, Schreiber e Wielinga (2002). Os autores procuram organizar uma biblioteca de teorias ontológicas reutilizáveis na área médica e classificam as ontologias em relação ao tipo de estrutura da conceitualização e em relação ao assunto da conceitualização. Distinguem as seguintes categorias:

- **Ontologias terminológicas:** especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos);
- **Ontologias de informação:** especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados);
- **Ontologias de modelagem do conhecimento:** especificam conceitualizações do conhecimento; têm uma estrutura interna mais rica e são refinadas para uso no domínio do conhecimento que descrevem;
- **Ontologias de aplicação:** contêm as definições que são necessárias para modelar o conhecimento em uma aplicação;
- **Ontologias de domínio:** expressam conceitualizações que são específicas para um domínio do conhecimento;
- **Ontologias genéricas:** são similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos;
- **Ontologias de representação:** explicam as conceitualizações que estão por trás dos

formalismos de representação do conhecimento.

Em outra abordagem, Mizoguchi, Vanwelkenhuysen e Ikeda (1995) também procuram definir tipos de ontologias com o objetivo de classificá-las. Os autores dividem as ontologias em:

- **Ontologias de domínio:** são reutilizáveis em um dado domínio e fornecem vocabulário sobre conceitos dentro desse domínio, sobre seus relacionamentos, sobre as atividades e sobre princípios que os governam;
- **Ontologias de tarefa:** fornecem um vocabulário sistematizado de termos usados para solucionar problemas, especificando tarefas que podem ou não estar no mesmo domínio;
- **Ontologias Gerais / Comuns:** incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções, etc.

Para Haav e Lubi (2001) ontologias “de alto nível”⁵⁶ descrevem conceitos gerais como espaço, tempo, matéria, objeto, evento, ação, etc, os quais são independentes do problema ou domínio. Por outro lado, 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.

De forma análoga Jasper e Uschold (1999) identificam quatro categorias principais de aplicações⁵⁷ de ontologias

- **Ontologia de autoria neutra:** um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas (reutilização de informações);
- **Ontologia como especificação:** uma ontologia para um domínio é criada e usada para desenvolvimento de softwares (documentação e manutenção);
- **Ontologia de acesso comum a informação:** quando a informação necessária é expressa por um vocabulário inacessível, a ontologia torna a informação inteligível, proporcionando conhecimento compartilhado dos termos, ou inter-relacionando grupos de termos (inter-operabilidade);

Mesmo sem um consenso, pode-se observar que os tipos apresentados guardam

⁵⁶ Fornecem noções gerais que podem estar relacionadas a todos os termos presentes em uma ontologia e são a idéia básica do meta-modelo apresentado na Capítulo 5.

⁵⁷ O autor explica que no contexto em estudo, entende-se por “aplicação” um sistema ou processo (em geral um programa de computador) que faz uso de uma ontologia.

semelhanças entre suas funções. Conhecidos os principais tipos, pode-se buscar ontologias já desenvolvidas e adequadas à tarefa a ser executada. A seção seguinte apresenta a importância da utilização das ontologias em sistemas computacionais. Em seguida são apresentados alguns exemplos de ontologias existentes e disponíveis na Internet, de vários tipos e domínios do conhecimento.

3.2.5 O papel das ontologias nos sistemas atuais

Os computadores não são mais simples dispositivos isolados, e se transformaram em “pontos de acesso” inseridos em redes, caracterizadas pelo intercâmbio constante de informações (FENSEL, 2001). Dessa forma, soluções que propiciam esse intercâmbio, ou inter-operabilidade, são essenciais no âmbito das tecnologias informacionais. Uschold e Gruninger (1996) também afirmam que a inter-operabilidade é uma importante aplicação para as ontologias.

Uma abordagem que tem recebido atenção é aquela que utiliza ontologias, atuando nas fontes de dados, para proporcionar melhor organização do conhecimento e, conseqüentemente, uma recuperação mais efetiva.

Como as ontologias possibilitam uma compreensão comum e compartilhada de um domínio do conhecimento, em que deve haver comunicação entre pessoas e sistemas, elas têm papel importante no intercâmbio de informações, pois proporcionam uma estrutura semântica às fontes de dados. Dessa forma, é possível a comunicação entre os agentes envolvidos nos processos (computadores ou pessoas), ao serem reduzidas diferenças conceituais ou terminologias.

O uso de ontologias em domínios como a modelagem de processos de negócios e arquiteturas de sistema, pode proporcionar um ambiente de integração entre as diferentes fontes disponíveis.

O uso das ontologias como ferramenta para inter-operabilidade é estudado intensivamente. Tipos de estruturas conceituais, ontologias e linguagens correlatas já são conhecidas e claramente categorizadas de acordo com suas funções. Uma grande variedade de tecnologias que se baseiam em ontologias já é utilizada por diversas comunidades científicas e industriais (esta diversidade está bem representada no Capítulo 4). Pelo apresentado até aqui, entende-se que as ontologias ocupam um lugar de destaque nos projetos de integração de fontes de dados. Cabe agora estudar de que forma podem contribuir.

3.2.5 Ontologias disponíveis na Internet

Existem ontologias desenvolvidas e disponíveis para uso e consulta. A construção de

ontologias sobre um domínio do conhecimento é uma tarefa pesada e portanto, deve-se aproveitar as ontologias existentes sempre que possível.

Apresentam-se a seguir alguns exemplos de repositórios de ontologias disponíveis na Internet. Não há a pretensão de listar todas, pois existem muitos outros recursos e bibliotecas de uso público, relativas a ontologias. As mais conhecidas são:

- *DAML*: banco com 192 ontologias, implementadas em *DAML+OIL*⁵⁸;
- *Ontolingua Server*: armazena aproximadamente 50 ontologias escritas em *Ontolingua*⁵⁹, além de informações e estatísticas.
- *Universal Repository*: armazena cerca de 50 ontologias para educadores e escolas, classificadas por áreas de conhecimento (história, computação, literatura, etc.)⁶⁰.

Existem também diversas ontologias desenvolvidas em domínios específicos do conhecimento. Por exemplo, Kohler e Schulze-Kremer (2002) avaliam ontologias existentes na área de biologia, apresentando uma tabela em que são listados e discutidos aspectos de algumas ontologias, vocabulários controlados e outras fontes de conhecimento estruturado.

Maedche (2000) acredita que a criação de ontologias a partir de dados existentes na Web é um típico trabalho de reengenharia e uma tarefa dispendiosa. Com relação ao reuso de ontologias, alguns pesquisadores são céticos e acreditam que é difícil usar as ontologias existentes em função dos fatores seguintes (GÓMEZ-PÉREZ, 1999):

- Não existe padronização das características das ontologias do ponto de vista do usuário;
- A formalização do conteúdo da ontologia difere em função do servidor em que está armazenada;
- Ontologias no mesmo servidor são em geral descritas com níveis de detalhes diferentes;
- Os *sites* da Web que usam a mesma organização lógica não apresentam informações relevantes sobre ontologias;
- As pesquisas por ontologias adequadas é tarefa difícil e que consome tempo.

Quando não é possível utilizar as ontologias pré-concebidas e existem recursos suficientes, a construção de ontologias pode ser uma possibilidade. Obviamente, muitas

⁵⁸ Trata-se de uma linguagem de marcação semântica para a Web, cuja especificação é definida pelo W3 Consortium (maiores detalhes em <<http://www.w3.org/TR/daml+oil-reference/>>).

⁵⁹ É um *software* que proporciona um ambiente distribuído e de trabalho em grupo para navegar, criar, editar modificar usar ontologias (<<http://www.ksl.stanford.edu/software/ontolingua/>>).

⁶⁰ Maiores detalhes em <<http://www.educanext.org/UNIVERSAL/>>.

ontologias devem ainda ser construídas. A sugestão é que, quando possível, seja aproveitado algum trabalho já existente. Ao se optar pela construção, várias metodologias, linguagens e ferramentas estão disponíveis. Algumas são citadas na seção seguinte, que enfatiza a construção de ontologias.

3.3 Breve pesquisa sobre a construção de ontologias

As técnicas para construção de ontologias têm se desenvolvido nos últimos anos, pois seu uso tem se difundido pelas iniciativas de especialistas de diversas áreas de conhecimento. Muitas comunidades científicas utilizam ontologias genéricas ou específicas, como por exemplo: a comunidade de Inteligência Artificial (no desenvolvimento de *sistemas baseados em conhecimento*), a comunidade de linguagens naturais (para caracterizar o sentido e o significado das palavras), a comunidade de bancos de dados (como esquema conceitual e para buscar interoperabilidade semântica entre bancos de dados heterogêneos), na recuperação da informação (abordagem semântica), e ainda, relacionadas a campos de estudo da Ciência da Informação, como os metadados (metadados semanticamente consistentes possibilitam recuperação de informação eficiente).

Existem situações em que é necessário desenvolver ontologias. Noy e Guinness (2001, p. 1) apresentam razões porque se deve desenvolver uma ontologia: “Para compartilhar a estrutura de informação comum entre pessoas ou softwares, para reutilizar do conhecimento do domínio, para tornar explícitos fatos consensuais, para separar um domínio do conhecimento do conhecimento operacional, para analisar um domínio”.

Construir uma ontologia inclui a definição das classes da ontologia, o arranjo das classes em uma hierarquia, a definição das propriedades ou restrições (conhecidos como *slots*) das classes e o preenchimento dos valores das propriedades (instâncias)⁶¹.

É importante ressaltar que não existe maneira única ou correta de modelar um domínio. O processo de construção é obrigatoriamente iterativo e os conceitos em uma ontologia podem estar inseridos em objetos ou relacionamentos, dentro do domínio de interesse (NOY; GUINNESS, 2001).

Ilustrando as diversas possibilidades existentes para a construção de ontologias, apresentam-se nas Seções 3.3.1, 3.3.2 e 3.3.3, metodologias, ferramentas e linguagens atualmente utilizadas nessa tarefa.

⁶¹ Uma ontologia será construída na Seção 5.3 utilizando essa sequência de atividades.

3.3.1 Metodologias para a construção de ontologias

Um grande número de metodologias têm sido desenvolvidas nos últimos anos no intuito de sistematizar a construção e manipulação de ontologias.

LÓPEZ (1999) analisa algumas metodologias (*Uschold & King, Gruninger & Fox, Bernaras et al, Methontology, Sensus*) em relação a norma *Institute of Electrical and Electronics Engineers (IEEE) Standard 1074-1995* e conclui que nenhuma das metodologias está madura se comparada ao padrão *IEEE*. Jones, Bench-Capon e Visser (1998), reconhecem que construir um modelo de domínio, ou ontologia, é um passo importante no desenvolvimento de *sistemas baseados em conhecimento*, mas concluem que a construção de ontologias ainda é muito mais artesanal do que científica.

Além das metodologias citadas acima, diversas outras estão disponíveis e cada uma delas pode ser mais adequada a uma situação em particular. Corcho, Fernández-López e Gomez-Pérez (2001) apresentam vários tipos de metodologias para construção e manipulação de ontologias, as quais são classificadas de acordo com suas características. As figuras do Anexo D sintetizam os comentários sobre cada metodologia apresentada, proporcionando uma visão geral de como funciona cada uma delas.

Após a escolha da metodologia para a construção da ontologia é importante verificar quais as ferramentas estão disponíveis para sua construção. Qualquer apoio na realização da tarefa pode representar ganhos significativos. A próxima seção apresenta diversas ferramentas, disponíveis para a construção de ontologias.

3.3.2 Ferramentas e linguagens para construção de ontologias

Como tarefa complementar pode-se estudar as referências a ferramentas (sem a pretensão de cobrir toda a literatura) disponíveis para a construção de ontologias, ferramentas para integração, ferramentas de anotação (utilizadas em páginas da *Web*), ferramentas de extração de ontologias e ferramentas para atender a consultas (*motores de inferência*).

Como não é objetivo desse trabalho a construção de complexas ontologias em um domínio também complexo, não se apresentam maiores detalhes das ferramentas. Para informações adicionais, pode-se consultar as referências apresentadas. Alguns exemplos de ferramentas para a construção de ontologias são apresentadas no Anexo D. As ferramentas para o auxílio à construção de ontologias apresentadas neste Anexo utilizam linguagens específicas para esta tarefa.

Algumas das linguagens utilizadas têm outras funções e são adaptadas para a tarefa,

enquanto outras foram desenvolvidas especificamente para trabalhar com ontologias. Muitas têm sido também desenvolvidas para trabalhar no ambiente da Internet, no contexto da *Web Semântica*. Para maiores informações, pode-se consultar as referências e os comentários também apresentados no Anexo D.

3.4 Considerações finais

Com um entendimento sobre o que seja uma ontologia, sobre sua construção e quais ferramentas estão disponíveis, é interessante que sejam apresentados casos práticos de sua utilização. Ontologias proporcionam a organização necessária para uma representação de maior qualidade, facilitando o acesso a fontes de dados. Para comprovar essa idéia, de forma prática, o próximo capítulo apresenta diversos aplicativos que usam ontologias e uma lista comentada de aplicativos que usam ontologias especificamente para integração.

CAPITULO 4 – ONTOLOGIAS E INTER-OPERABILIDADE

4.1 Introdução

A Internet tem possibilitado acesso a fontes de dados numa escala sem precedentes. Existem benefícios óbvios, mas também problemas relacionados à dificuldade em acessar as fontes de dados de forma eficiente. A função de organização da informação, antes restrita a especialistas (bibliotecários) hoje deve ser exercida em muitas outras situações e por outras pessoas. O uso de ontologias pode facilitar esta tarefa, à medida que procura classificar e compartilhar algum consenso em determinado domínio de interesse.

Ontologias possibilitam uma compreensão comum e compartilhada de um domínio do conhecimento, em que deve haver comunicação entre pessoas e aplicativos. Desse ponto de vista, trata-se de um tópico relevante no contexto de estudos sobre inter-operabilidade, pois determinar conceitos e relações em domínios específicos é parte importante de qualquer projeto de integração.

Conforme citado anteriormente, a extração de informações de fontes de dados e sua integração em uma empresa pode envolver sistemas de suporte a decisão, *warehouses*, bibliotecas digitais, sistemas de recuperação da informação, correio eletrônico e sistemas distribuídos interligados pela Internet ou por redes privadas. Todos esses sistemas, e, em última instância o próprio usuário, podem se beneficiar da consulta à base de dados mais organizadas.

Em função da necessidade de interoperar, inerente aos ambientes atuais, muitos esquemas de metadados têm sido propostos nos últimos anos. Os metadados podem ser organizados em ontologias que classificam suas categorias e possibilitam mais informações sobre um determinado domínio.

Nas Seções 4.2 e 4.3, procura-se mostrar como as ontologias podem ser utilizadas para a organização do conhecimento. São apresentadas também aplicações baseadas em ontologias e aplicações que usam ontologias para a inter-operabilidade.

4.2 Aplicações baseadas em ontologias

Além de outras utilizações, citadas no Capítulo 3, as ontologias são uma abordagem possível para se tratarem problemas de integração. Segundo Pacheco e Kern (2001, p.58), “[...] a construção de ontologias comuns tem sido proposta como abordagem promissora para a inter-operabilidade de sistemas”.

Muitas aplicações que envolvem ontologias tratam a questão da inter-operabilidade e

estão presentes em diversos campos do conhecimento. Um esquema gráfico que identifica formas de aplicação é apresentado na Figura 9:

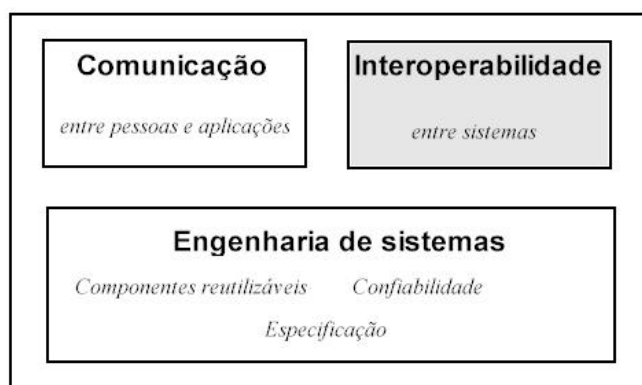


Figura 9 – Formas de aplicação de ontologias (USCHOLD; GRÜNINGER, 1996)

Conforme a Figura 9, uma das formas de utilizar ontologias é na inter-operabilidade entre sistemas. Esses sistemas podem consultar diversas fontes de dados e sua integração será mais eficiente através da organização, formalização e tradução entre os diferentes significados ou conceitos presentes. Gómez-Pérez (2001) corrobora com essa idéia ao afirmar que as ontologias podem ser utilizadas na comunicação entre pessoas, organizações e na inter-operabilidade entre sistemas (tradução de métodos de modelagem, paradigmas, linguagens e ferramentas de *software*).

Aplicações que utilizam ontologias são encontradas em diversos campos do conhecimento. Algumas dos tipos de aplicações mais conhecidos em domínios diversos são citadas em seguida e listadas no Anexo E. Não se trata de uma lista completa e nem se tem a pretensão de cobrir toda a literatura sobre o assunto. As aplicações relativas à integração de fontes de dados que utilizam ontologias, de maior interesse para este trabalho, são tratadas com mais detalhes na Seção 4.3.

Encontram-se projetos em andamento, relacionados a ontologias, em gestão do conhecimento, comércio eletrônico, processamento de linguagens naturais⁶², recuperação da informação na Web e em educação (CORCHO; FERNÁNDEZ-LÓPEZ; GOMEZ-PÉREZ, 2001). Existem também projetos comerciais para a recuperação da informação baseados em ontologias (HAAV; LUBI, 2001). As figuras do Anexo D sintetizam os comentários sobre cada projeto ou ferramenta apresentado.

Alguns projetos abordam a integração das próprias ontologias, entre si. Por exemplo, Wache et al. (2001) apresentam a integração de ontologias sobre dois aspectos: integração de

⁶² Manipulação, análise e compreensão da linguagem natural pelo computador.

ontologias às fontes de dados e integração inter-ontologias. O segundo aspecto é conhecido como “mapeamento entre ontologias” e mais de uma ontologia é usada para descrever a informação.

Uma abordagem comum para o problema é possibilitar ao usuário a definição do mapeamento, o que permite grande flexibilidade. Entretanto, não se consegue preservar a semântica original, visto que o usuário pode alterá-la. Como exemplo para essa abordagem é citado o *KRAFT*⁶³.

Outra possibilidade é estabelecer uma correlação mínima entre a semântica das ontologias, de forma que se possa efetuar o mapeamento. Um exemplo é o *OBSERVER*⁶⁴, que utiliza técnicas da lingüística e relações léxicas (por exemplo, sinônimos) para que o mapeamento possa ser executado, mantendo-se pelo menos parte da semântica.

Existe também a possibilidade de se permanecer atrelado a um formalismo, de mais alto nível, e mapear as ontologias para aquela representação. Isso pode ser feito com a idéia de herança a partir dos conceitos da ontologia de alto nível. Um exemplo de abordagem desse tipo, cujo formalismo é baseado em lógica descritiva, é o *DWQ*⁶⁵. A lógica descritiva e suas variações são elementos importantes na definição do meta-modelo desenvolvido neste trabalho.

Torna-se claro, a partir dos exemplos acima, que as ontologias encontram aplicação em diversos campos do conhecimento. Em particular, são estudadas atualmente como solução para problemas de integração de fontes de dados heterogêneas. Com relação às pesquisas na área, as tendências apontam para estudos de critérios de mensuração, bibliotecas de ontologias, uso intensivo de metodologias para desenvolvimento, dentre outros Gómez-Pérez (2001).

Em conformidade com as tendências apresentadas acima, várias iniciativas têm se desenvolvido para uso de ontologias em projetos de integração. Na seção seguinte são apresentados diversos projetos que aplicam ontologias para tratar questões especificamente de inter-operabilidade.

4.3 Trabalhos relacionados

4.3.1 Aplicações baseadas em ontologias específicas para a integração

Organizar, categorizar, prover significado são ações essenciais para se obter inter-operabilidade, tarefas que podem ser executadas por intermédio das ontologias. A Seção 4.2 proporciona uma visão geral das possibilidades de uso das ontologias em sistemas com objetivos

⁶³ Maiores detalhes na Seção 4.3.7.

⁶⁴ Maiores detalhes na Seção 4.3.5.

⁶⁵ Maiores detalhes na Seção 4.3.9.

diversos. Destacam-se, dentre essas aplicações, aquelas relativas às ontologias utilizadas para integração de fontes de dados, as quais são ilustradas nessa Seção. Algumas dessas aplicações serão apresentadas com mais detalhe em seguida visto que executam trabalho análogo ao que se pretende conduzir para integração no âmbito do meta-modelo. Esquemas das arquiteturas de cada sistema são apresentados no Anexo F.

4.3.2 TSIMMIS

O *Stanford/IBM Manager of Multiple Information Sources (TSIMMIS)* é um projeto pioneiro na integração de dados e concebido pela comunidade de banco de dados. O objetivo do *TSIMMIS* é desenvolver ferramentas para a integração de fontes de dados heterogêneos, que podem incluir dados estruturados ou não.

Segundo Chawathe et al. (1994), para cada fonte de informação considerada, o sistema aloca um *wrapper*⁶⁶ que converte os objetos representativos dos dados em um modelo comum. Este modelo auto-descritivo (todos os objetos têm rótulos que descrevem seu significado) é chamado *Objetc Exchange Model (OEM)*. O *wrapper* converte as consultas, executadas no formato do modelo comum, em pedidos que a fonte pode executar. Os dados retornam da fonte convertidos em dados do modelo comum.

Um mediador processa respostas antes de encaminhá-las ao usuário, convertendo dados para o formato comum e eliminando redundâncias. O mediador trabalha de forma independente das fontes que vai usar, pois quando passa objetos para seus clientes, adiciona o nome da fonte de forma que o cliente possa interpretar cada objeto (WIEDERHOLD, 1992). Assim os usuários ou componentes de software não precisam ter uma visão global de toda a informação manipulada pelo sistema. *Wrappers* e mediadores recebem como entrada consultas em uma linguagem específica⁶⁷ e retornam objetos *OEM*. Os usuários finais podem acessar a informação escrevendo aplicações que recebem objetos *OEM* ou usando um navegador genérico.

O gerenciamento de restrições é feito pelo *Constraint Manager*, que especifica semanticamente a necessidade de consistência sobre os dados armazenados.

A extração é feita a partir de fontes de dados não estruturadas. Um componente (*classifier/extractor*) classifica automaticamente objetos das fontes (*e-mail*, texto, imagens, etc) e extrai suas propriedades-chave (por exemplo, autor, data de criação, etc), baseando-se na identificação de padrões dos objetos. A informação coletada pode ser enviada para qualquer outro

⁶⁶ Corresponde ao *translator* na arquitetura, vide figura do Anexo G.

⁶⁷ No caso, a OEM-QL, *OEM-Query Language*.

componente do sistema *TSIMMIS*.

Para Mena et al. (1996), o projeto *TISMMIS* trabalha a partir da geração semi-automática de *wrappers* e mediadores que fazem correspondência da informação com o modelo *OEM* para dados estruturados ou não. Referências de projetos que adotam abordagens similares são: projeto *HERMES* (SUBRAHMANYAN et al., 1992) e projeto *DISCO* (TOMASIC; RASCHID; VALDURIEZ, 1995).

O projeto *TSIMMIS* não utiliza ontologias para descrever os repositórios de dados, mas um mapeamento inter-ontologias identifica semanticamente a correspondência entre termos de diferentes ontologias (ou seja, quais termos são semanticamente iguais ou similares). Conforme (WACHE et al., 2001), o projeto *TSIMMIS* correlaciona diretamente a ontologia com o esquema do banco de dados. É produzida uma cópia da estrutura do banco de dados, a qual é codificada em uma linguagem que possibilite a formulação automática de inferências.

4.3.3 *SIMS*

Segundo Arens, Hsu e Knoblock (1996), o *SIMS* é um mediador⁶⁸ que acessa e integra diversas fontes de dados. As consultas são expressas em uma linguagem comum, independente das várias linguagens de consulta e da localização das fontes. O *SIMS* determina quais fontes usar, como obter a informação desejada, como e onde armazenar temporariamente dados, como manipulá-los e como recuperá-los.

O mediador contém um modelo do domínio (base de conhecimento hierárquica e terminológica) e um modelo de todas as fontes de informação (ambos expressos em linguagem *LOOM*⁶⁹). A modelagem descreve o relacionamento entre classes, subclasses e superclasses, a função de cada classe, o seu conteúdo e a integração desses modelos com o modelo do domínio.

As modelagens executadas pelo sistema são, assim:

- Modelagem do domínio: onde o sistema traduz a consulta executada no domínio em um conjunto de consultas para as fontes de informação (as quais podem ser bancos de dados, bases de conhecimento, etc).
- Modelagem de fontes de informação: em que todos os conceitos e relacionamentos no modelo da fonte de dados são traduzidos para conceitos e relações no modelo domínio.

A fonte de dados é selecionada para atender a uma consulta expressa em termos do

⁶⁸ Sempre que citado, o termo tem mesma a conotação apresentada no item 4.3.2.

⁶⁹ Linguagem e ambiente para a construção de aplicações em inteligência artificial, desenvolvido pela University of Southern California .

modelo do domínio. O sistema aplica um conjunto de operadores para transformar conceitos do domínio em outros, que podem ser recuperados diretamente da fonte de dados (possui operadores como *Select-information-Source*, *Generalize-concept*, *Specialize-concept* e *Decompose-relation*, dentre outros).

Uma vez que o sistema reformulou as consultas para usar termos dos modelos de dados das fontes, gera-se um plano de consultas para recuperação e processamento dos dados, o qual especifica quais operações e em que ordem devem ser executadas. Não é necessário conversão entre linguagens.

A racionalização das consultas é executada com o objetivo de buscar a consulta menos dispendiosa dentre as semanticamente equivalentes. A transformação de uma consulta em outra é feita através de inferências lógicas usando-se abstrações que descrevem os bancos de dados com um conjunto de formulas de lógica de 1ª ordem⁷⁰.

Segundo Wache et al. (2001), o *SIMS* trabalha com uma única ontologia que fornece um vocabulário compartilhado para especificação da semântica (o *SIMS* inclui uma base terminológica). Com relação à conexão com as fontes de dados (esquema do banco de dados) é produzida uma cópia da estrutura a qual é codificada em linguagem que possibilite deduções automáticas (inferências). A integração é executada sobre a cópia, de forma a manter a base original inalterada.

4.3.4 COIN

O *Context Interchange Project (COIN)* lida com a integração de fontes de dados distribuídas e heterogêneas, adotando uma arquitetura de mediadores cujo principal componente é o MINT. Trata-se de um mediador que proporciona consultas a múltiplas fontes e tem a capacidade de detectar e resolver conflitos semânticos. *Wrappers* fazem a ligação física e lógica entre as fontes de dados e o mediador (BRESSAN, 1997).

O projeto *COIN* possui um protótipo cujos componentes executam três grupos diferentes de processos:

- Processos-Cliente: proporcionam interação entre os pedidos ao banco de dados e um mediador (*Context Mediator*);
- Processos-Servidor: *wrappers* permitem consultas a documentos semi-estruturados na Web (como se fosse bancos de dados comuns); e, *gateways* do banco de dados, que conectam

⁷⁰ Linguagem que descreve verdades lógicas através de fórmulas matemáticas e utiliza operadores chamados “conectivas”.

fisicamente os bancos de dados a uma rede;

- Processos-Mediadores: componentes do sistema que fornecem serviços coletivos de mediação; incluem o *Context Mediator* (reescreve uma consulta do usuário em uma consulta mediada), o *Optimizer* (planeja a avaliação das consultas baseado em consultas mediadas) e o *Executioner* (executa o plano enviando sub-consultas para os processos servidores, operando os resultados intermediários e retornando as respostas finais).

O *COIN* adota uma abordagem híbrida, ou seja, cada fonte de dados é descrita por sua própria ontologia. Para que as ontologias locais (descritas como um vetor a partir de descrições locais chamadas “contextos”) sejam comparáveis, elas são construídas a partir de um vocabulário global compartilhado. Este vocabulário contém termos básicos (primitivas), os quais são combinados em ontologias locais para descrever semânticas mais complexas (WACHE et al., 2001).

O *COIN* adota uma representação “*frame-based*”⁷¹ e usa *F-logic*⁷² aliada a uma sintaxe definida pelo próprio sistema.

4.3.5 OBSERVER

O conteúdo das fontes de dados é descrito por conceitos presentes em ontologias que representam um domínio do conhecimento. Para Mena et al. (1996), o problema diz respeito ao compartilhamento do vocabulário, ou seja, lidar com diferentes termos ou conceitos para descrever informações similares. O *Ontology Based System Enhanced with Relationships for Vocabulary heterogeneity Resolution (OBSERVER)* permite a representação de ontologias por relacionamentos “inter-ontologias” e correlaciona os termos entre elas. As consultas do usuário são reescritas usando esses relacionamentos, o que resulta em traduções entre as ontologias.

Em geral, existem diferenças entre as ontologias pois são desenvolvidas por organizações diferentes e de forma independente. Ontologias diferentes com vocabulários diferentes podem atender melhor às necessidades dos usuários. O *OBSERVER* tem a capacidade de lidar com múltiplas ontologias, mas fica assim também responsável pela tarefa (muitas vezes, complexa) de manter consistência e eficiência. Ao se usar a abordagem múltipla, as ontologias devem estar interligadas por relacionamentos inter-ontologias e organizadas em grupos que correspondem a áreas do conhecimento. Como alguns grupos são mais gerais que outros, podem

⁷¹ Baseada em “*frames*”, estruturas de dados que contém variáveis pertencentes a um determinado escopo. (Minsky, 1981).

⁷² Linguagem de programação orientada a objetos e sistema e banco de dados dedutivo, ou seja, que combina um banco tradicional (fatos), uma base de conhecimento (regras) e uma máquina de inferência que deriva informações desses fatos e regras.

ser organizadas hierarquias.

No *OBSERVER*, as ontologias são descritas por um sistema de lógica descritiva chamado *CLASSIC* (outros exemplos são *BACK* (LUCK et al., 1987 citado por MENA et al., 1996), *LOOM* (KRIS, BAADER; HOLLUNDER, 1991⁷³ citados por MENA et al., 1996)). As ontologias são reutilizadas após sua representação pelo *CLASSIC*. Foram criadas usando linguagens de representação diferentes, mas são construídas em lógica descritiva, tentando-se manter a mesma semântica original. Usam-se estas ontologias para descrever repositórios de dados.

A arquitetura do *OBSERVER* espelha a necessidade de se criar um sistema independente do número de ontologias e capaz de lidar com problemas de heterogeneidade. Com relação ao vocabulário, usa-se um repositório compartilhado contendo as relações inter-ontologias, chamado *Interontology Relationship Manager (IRM)*. O *IRM* pode ser visto como um “catálogo da semântica” do sistema.

O *OBSERVER* aborda o problema de integração através de múltiplas ontologias (WACHE et al., 2001). A vantagem dessa abordagem é que não é necessária uma ontologia global e as ontologias de cada fonte de dados podem ser desenvolvidas sem preocupação com as outras fontes (ou com suas respectivas ontologias). Por outro lado, a ausência de um vocabulário comum torna difícil a comparação de diferentes ontologias de fontes de dados. Para solucionar este problema, é necessário um outro formalismo de representação definindo um relação “inter-ontologias”.

4.3.6 INFOSLEUTH

Conforme Mena et al. (1996), o projeto *INFOSLEUTH* utiliza uma rede de agentes para coleta e análise de informações em um sistema global (por exemplo, a Internet). Recupera e processa dados em fontes sujeitas a mudanças freqüentes. Integra diversas tecnologias como agentes, ontologias e computação distribuída, para mediação da interoperação de dados e serviços em um ambiente dinâmico.

Os agentes comunicam-se usando uma ontologia especial e usam ontologias específicas para captura de dados. A consulta é expressa em conceitos ontológicos os quais são interpretados por agentes. Estes então, se comunicam com outros e retornam uma resposta ao usuário (HWANG, 1999). O *INFOSLEUTH* trata uma fonte de dados no nível semântico e utiliza ontologias para capturar dados de diversas fontes.

⁷³ BAADER, F., HOLLUNDER, B. KRIS: Knowledge representation and inference system. *SIGART Bulletin*, v. 2, n. 3, p. 8-14, 1991.

No *INFOSLEUTH* os agentes de comunicação utilizam KQML⁷⁴. Os usuários especificam consultas nas ontologias via uma interface baseada em *applets*⁷⁵. A linguagem de representação do conhecimento KIF e a linguagem de consulta de banco de dados SQL são utilizadas internamente para representar consultas em ontologias específicas.

Os elementos da arquitetura do *INFOSLEUTH* são (BAYARDO et al., 1997):

- **Agente do usuário:** representante inteligente do usuário que usa o seu conhecimento das ontologias para auxiliar o usuário na formulação de consultas e obtenção dos resultados;
- **Agente da ontologia:** torna disponível o conhecimento das ontologias e fornece respostas a consultas sobre elas;
- **Agente de intermediação:** recebe e armazena mensagens de todos os agentes do sistema sobre suas capacidades; baseado nessas informações, responde a consultas de agentes, direcionando pedidos específicos;
- **Agente de recursos:** correlaciona a ontologia comum para o esquema do banco de dados e sua linguagem nativa, executando pedidos específicos para aquele recurso;
- **Agente de análise de dados:** coordena a execução de sub-tarefas de coleta de informação de alto nível, necessária para atendimento às consultas;
- **Agente monitor:** acompanha as interações dos agentes e os passos de execução das tarefas; fornece uma interface visual para exibir a execução.

O *INFOSLEUTH* aceita ontologias de mais de um domínio, mas não possibilita a tradução de consultas através das diversas ontologias.

4.3.7 KRAFT

O *KRAFT* é uma das linguagens usadas em sistemas de integração baseados em ontologias, conhecidas como linguagens de representação *frame-based* (WACHE et al., 2001). O *KRAFT* adota a abordagem conhecida como *fusão de conhecimento*⁷⁶, através do qual as organizações podem combinar o conhecimento de fontes diversas de maneira altamente dinâmica.

⁷⁴ *Knowledge Query and Manipulation Language* – Linguagem e protocolo, muito utilizado na área de Inteligência Artificial, para troca de informações e conhecimento. Pode ser usado com linguagem de consulta para sistemas e agentes.

⁷⁵ Um programa escrito em linguagem Java que pode ser distribuído como um anexo a uma página da Web e ser executado por um navegador que execute Java.

⁷⁶ Processamento de informações que trata da associação, correlação e combinação de dados de uma ou mais fontes, com o objetivo de obter uma avaliação mais completa e precisa da situação. O processo é caracterizado pelo refinamento contínuo das estimativas e pela avaliação da necessidade de fontes adicionais ou modificação do próprio processo, para se alcançar resultados mais satisfatórios.

Os sistemas submetidos a fusão usam algum tipo de esquema de integração para conseguir uma representação comum no sistema distribuído. As fontes de dados individuais são atribuídas a seus esquemas locais e ao esquema de integração, para que os dados possam ser combinados (PREECE; HUI, 1999).

Para combinar informações, as instâncias de dados devem usar o conhecimento associado ao seu contexto, além de saber como elas são interpretadas e usadas. Este tipo de conhecimento pode ser expresso na forma de *restrições*⁷⁷.

Em um sistema que usa a fusão, as instâncias de dados e respectivas restrições associadas são atribuídas a uma representação comum permitindo a fusão. É necessário uma linguagem comum para representação das instâncias e restrições e um grupo comum de definições de terminologia do domínio de conhecimento, ou seja, uma ontologia compartilhada. As restrições são expressas em *CoLan*⁷⁸, que serve como uma linguagem única (de restrições) e é relativamente legível.

O conhecimento presente nos recursos individuais precisa ser transformado em uma linguagem comum, nos termos da ontologia compartilhada, antes que possa ser combinado.

Os componentes de processamento de conhecimento da arquitetura do *KRAFT* são concebidos como *agentes de softwares*. O projeto da arquitetura *KRAFT* é consistente como diversos padrões de agentes, como por exemplo o KQML.

Existem três tipos de agentes na arquitetura do *KRAFT*: os *wrappers*, mediadores e facilitadores. Todos eles, de alguma forma processam conhecimento:

- **Wrappers:** agentes que funcionam como “pontes” para fontes de dados externas, em geral bancos de dados e sistemas baseados em conhecimento; também podem proporcionar a interface entre um sistema legado e o *KRAFT*;
- **Mediadores:** agentes de processamento do conhecimento internos do sistema; cada mediador agrega valor ao conhecimento obtido por outros agentes a partir de suas tarefas típicas como seleção, ordenação e fusão do conhecimento obtido por outros agentes;
- **Facilitadores:** agentes que possibilitam “familiaridade” entre outros agentes para que possa haver comunicação entre eles;

Os agentes do *KRAFT* se comunicam utilizando dois protocolos:

⁷⁷ Por exemplo, em um sistema distribuído para o projeto de computadores pessoais, uma restrição pode ser “Windows NT requer no mínimo 32 Mb de memória RAM”.

⁷⁸ *Constraint Language*, linguagem baseada em restrições.

- **CCQL – *Constraint Command and Query Language***: subgrupo do KQML, que funciona como o protocolo externo;
- **CIF – *Constraint Interchange Format***: subgrupo da linguagem *CoLan* que expressa o conteúdo da mensagem.

4.3.8 **PICSEL**

O *PICSEL* consiste de um mediador e é projetado com duas partes principais: uma *máquina de consultas* e *bases de conhecimento* específicas, as quais contêm modelos do domínio e descrições do conteúdo das fontes de informação acessíveis (GOASDOUÉ; REYNAUD, 1999).

Em um servidor de informações, existe uma base de conhecimento para modelar o domínio e uma base de conhecimento que descreve o conteúdo de cada fonte. O modelo de domínio contém todo o vocabulário usado para responder às consultas. O motor de consultas acessa as fontes para responder às consultas dos usuários. O conteúdo das fontes de informação é representado no mesmo formalismo lógico das consultas dos usuários e dos dados do domínio. *Wrappers* trabalham como módulos especializados que se referem a cada modelo de dados. Quando uma fonte é um banco de dados relacional, *wrappers* traduzem a consulta em termos das relações das fontes para uma forma relacional.

A maioria dos sistemas mediadores tem em comum o uso de um formalismo lógico para descrever tanto o modelo do domínio quanto o conteúdo das fontes de informação. Esses sistemas diferem, primeiro, pelo formalismo lógico usado, e segundo, pela maneira como expressam o conteúdo das fontes de informação, em relação ao modelo do domínio. A lógica descritiva também é utilizada na integração para modelar descrições complexas de dados.

No *PICSEL*, usa-se a linguagem *CARIN*⁷⁹ para representar o domínio escolhido e o conteúdo das fontes de informação que estão disponíveis. A *CARIN* contém dois componentes: um grupo de regras e um grupo de declarações de lógica descritiva que são definições sobre conceitos e papéis no domínio (relações binárias e unárias).

Durante o processo de integração várias correlações devem ser mapeadas de um esquema global para o esquema da fonte local. Uma sub-consulta está correta em relação à consulta global, se a sub-consulta local fornece uma parte das respostas solicitadas (isto é, as sub-consultas devem estar contidas na consulta global). Como uma ontologia possui uma especificação completa da conceitualização (GRUBER, 1993b), as correlações podem ser

⁷⁹

Vide figura do Anexo D.

validadas pela ontologia. Os conceitos da ontologia correspondentes às sub-consultas locais estão contidos nos conceitos da ontologia relacionados à consulta global.

4.3.9 DWQ

Os sistemas de integração baseados em ontologias são diferentes quanto à análise do conteúdo das fontes. Podem ser identificadas três formas de emprego das ontologias para essa tarefa: uma ontologia única (usa-se uma única ontologia global e um vocabulário compartilhado para a especificação da semântica); diversas ontologias (cada fonte de informação é descrita por sua própria ontologia); e, híbrida (possui diversas ontologias para a semântica de cada fonte, mas também possui um vocabulário global compartilhado, de forma que as ontologias locais possam ser comparadas) (WACHE et al., 2001). O *Data Warehouse Quality (DWQ)* (CALVANESE et al., 1998) parece ser o único sistema em que as três abordagens podem ser implementadas.

No DWQ, além da função de análise do conteúdo, a ontologia também tem a tarefa de descrever o processo de integração⁸⁰. O processo de integração utiliza-se de representação por lógica descritiva. Assume-se que cada fonte é uma coleção de tabelas relacionais e cada uma delas é descrita em termos de sua ontologia com o auxílio de consultas. Uma consulta global e sua decomposição em sub-consultas é relacionada a conceitos ontológicos. As sub-consultas estão corretas, isto é, estão contidas na consulta global, se seus conceitos ontológicos estão contidos nos conceitos da ontologia global.

O projeto DWQ proporciona a base semântica que permite aos projetistas de grandes bancos de dados interligar seus complexos modelos e utilizar estruturas de dados cada vez mais ricas, de maneira sistemática, melhorando o projeto, a operação e a evolução das bases de dados, ou seja, proporcionando qualidade (JARKE; VASSILIOU, 1997).

O projeto DWQ fornece um modelo de arquitetura de referência abrangendo o projeto, a configuração, a operação, manutenção e a evolução de *warehouses*. Os elementos da arquitetura⁸¹ são:

- **Fontes:** qualquer repositório cujo conteúdo pode ser utilizado em uma base de dados;
- **Wrappers:** descrevem as fontes de dados no formato da base genérica;
- **Bancos de dados destino:** grandes fontes de dados (por exemplo: *warehouses* e *data-marts*⁸²);

⁸⁰ Essa característica será utilizada adiante para descrição da integração do meta-modelo proposto.

⁸¹ Esboço disponível no Anexo F.

⁸² A idéia e a tecnologia são as mesmas do *data-warehouse*, mas o banco de dados só é implementado em um setor específico da organização.

- **Banco de dados para metadados:** repositório para informação sobre outros componentes (por ex. esquema da fonte de dados);
- Agentes para administração;
- Clientes para apresentação dos dados.

O Projeto *DWQ* é baseado em uma arquitetura composta por diversos módulos e tecnologias. Torna disponíveis ferramentas que atendem às rigorosas exigências de projeto e às características de *warehouses*. Essas ferramentas têm por objetivo proporcionar qualidade ao processamento dos dados e são orientadas por uma abordagem rica em semântica.

4.3.10 ONTOBROKER

Além da lógica descritiva, o outro grupo principal de linguagens usado em sistemas de integração baseados em ontologias são as linguagens de representação clássicas, conhecidas com *frame-based*. O *ONTOBROKER* faz uso direto da *F-Logic* com uma sintaxe pré-definida.

Existem diversas possibilidades conexão entre a ontologia outras partes do sistema. O *ONTOBROKER* utiliza a abordagem de “meta-anotações” que adicionam informações semânticas às fontes de informação. Em geral, meta-anotações consistem de marcações especiais em linguagens conhecidas da Internet (extensões), ou do desenvolvimento de uma linguagem específica. Essa abordagem tem sido muito utilizada para integração fontes de dados na Internet, onde anotação é uma forma natural de proporcionar semântica. Outro projeto que utiliza essa técnica é o *SHOE*, que será apresentado na seção seguinte. O *ONTOBROKER* utiliza as anotações também para evitar a redundância de dados.

O *ONTOBROKER* é uma arquitetura que projeta e implementa ferramentas necessárias ao uso de ontologias para realizar consultas na Internet. Essa arquitetura é composta por três elementos principais: uma interface de consulta para usuários, uma máquina de inferência (usada para se obter respostas) e um agente inteligente (utilizado para coletar dados da Web) (FENSEL et al., 1998).

Oferece também uma linguagem de representação para a criação de ontologias, da qual, um subgrupo é utilizado para elaborar consultas (ou seja, uma linguagem de consultas é definida). Também possui uma “linguagem de anotação” para permitir ao fornecedor de informações “marcar” semanticamente documentos da Internet com informações ontológicas.

4.3.11 SHOE

Assim como no *ONTOBROKER*, o *Simple HTML Ontology Extensions (SHOE)* também utiliza meta-anotações, na forma definida na seção anterior, que adicionam semântica às

fontes de dados. Conforme citado, a anotação é uma forma natural de proporcionar semântica na Internet. Uma abordagem similar (*Formal Ontology Markup Language – FOML*) é apresentada anteriormente no Anexo D por Ogata (2001).

Ao invés de tentar apreender o conhecimento a partir de documentos *HTML* existentes, a opção é dar aos autores a habilidade de inserir informações diretamente na página (metadados), tornando a tarefa de recuperação da informação mais eficiente. Uma possibilidade é desenvolver um subgrupo do *HTML* que adicione conhecimento à sintaxe da marcação. As informações adicionadas não são simples palavras-chave. Mais do que isso, são “marcações semânticas” definidas em um grupo de atributos e relacionamentos (ou seja, uma ontologia).

O *SHOE* consiste de um grupo de extensões à linguagem *HTML* que permite que os autores de páginas da *Web* façam anotações, proporcionando correlações semânticas (LUKE et al., 1997). Estas anotações são expressas em conhecimento ontológico, que pode ser gerado pelo uso direto ou por uma extensão de ontologia disponível na *Web*. Isso torna possível fazer consultas eficientes, via agentes inteligentes, e não apenas por palavras-chave (como é feito atualmente pelos mecanismos de busca).

O *SHOE* proporciona recuperação da informação mais eficiente na *Web*, ao tornar mais fácil o acesso a dados. Possibilita ainda:

- Definir ontologias usando *HTML*, as quais representam esquemas de classificação e regras de modelos entidade-relacionamento;
- Criar novas ontologias que complementam ontologias existentes;
- Declarar entidades para documentos e subseções de documentos;
- Declarar relacionamentos entre entidades;
- Classificar entidades sobre um esquema de classificação do tipo “*is-a*”.

Luke et al. (1997) apresentam ainda comparações das habilidades do *SHOE* em relação ao esquema de *metatags*⁸³ do *HTML* e com o modelo entidade-relacionamento de banco de dados.

4.3.12 BUSTER

O *BUSTER* é um exemplo de sistema de integração baseado em ontologias que apresenta como característica a utilização de uma abordagem híbrida de ontologias para descrever

⁸³ Metadados inseridos em página *HTML* a partir de tags definidos na linguagem.

a semântica das fontes de informação. Na abordagem híbrida, já citada anteriormente nas seções 4.3.4 (*COIN*) e 4.3.9 (*DWQ*), a semântica de cada fonte é descrita por sua própria ontologia, a qual é construídas a partir de um vocabulário global compartilhado.

No *BUSTER*, o vocabulário compartilhado é uma ontologia (global), que define as faixas de valor de atributos e seus conceitos (WACHE et al., 2001). Uma ontologia-fonte é um refinamento (parcial) da ontologia geral que restringe a faixa de valores de alguns atributos. Como as ontologias-fonte usam apenas o vocabulário da ontologia geral, as duas permanecem comparáveis.

A conexão entre a ontologia e o sistema pode ser feita de várias maneiras. O *BUSTER* usa a ontologia para descrever melhor os termos de um banco de dados ou de seu esquema.

O sistema *BUSTER* parte do princípio que a análise e transformação do contexto do conhecimento é necessária para alcançar inter-operabilidade em nível semântico. Aborda o problema de preservar o significado de atributos únicos no banco de dados, visto que terminologias diferentes podem ser usadas (STUCKENSCHMIDT; WACHE, 2000).

Regras de integração orientam a correlação das fontes de dados para uma visão integrada fornecida pelo componente mediador. As regras se baseiam em uma descrição de objetos de cada fonte de dados. Assim, cada objeto, tabela de banco relacional ou um simples número, será “encapsulado” por um objeto padrão. Este objeto é um predicado, conforme definido na lógica, com campos que contem variáveis e descrevem instâncias encontradas no banco de dados.

4.4 Quadro sinótico

Finalizando o Capítulo 4, apresenta-se a Figura 10, com resumo sinótico dos diversos projetos apresentados:

Projeto	Princípio de Funcionamento	Papel da ontologia	Representação da ontologia	Comentários
<i>TSIMMIS</i>	Arquitetura de <i>wrappers</i> e mediadores	Mapeamento inter-ontologias identifica termos equivalentes entre diferentes ontologias.		Usa um modelo próprio (OEM) e uma linguagem de consulta (OEM-QL).
<i>SIMS</i>	Arquitetura de mediadores	Uma ontologia única fornece vocabulário compartilhado para especificação semântica.	Lógica descritiva (LOOM)	Possui operadores para transformar conceitos do domínio em conceitos da fonte de dados.
<i>COIN</i>	Arquitetura de <i>wrappers</i> e mediadores	A semântica de cada fonte é descrita por sua própria ontologia, a qual é construída a partir de um vocabulário comum compartilhado.	Linguagem <i>frame-based</i> (F-Logic)	Solução de conflitos semânticos é feita por um componente mediador especializado (MINT).
<i>OBSERVER</i>	Relações inter-ontologias usam princípios lingüísticos	Ontologias diferentes representam, de forma independente, as fontes de dados.	Lógica descritiva (CLASSIC)	O vocabulário é definido por um repositório compartilhado (IRM), que armazena as relações inter-ontologias.
<i>INFOSLEUTH</i>	Arquitetura de agentes constrói ontologias a partir de bancos de dados textuais	Ontologias específicas para as fontes de dados se relacionam através de agentes com uma ontologia especial que retorna as consultas.	Linguagem de representação do conhecimento (KIF)	Usa o KQML para comunicação dos agentes, SQL e KIF para representação das consultas.
<i>KRAFT</i>	Arquitetura de mediadores e <i>wrappers</i> (representados por agentes) executam “fusão do conhecimento”,	Ontologia compartilhada caracteriza totalmente a área do conhecimento referente ao problema.	Linguagem <i>frame-based</i>	Fusão conhecimento é um processo baseado em restrições (expressas em <i>CoLan</i>); agentes se comunicam em KQML e CIF.
<i>PICSEL</i>	Arquitetura de <i>wrappers</i> e mediadores que consiste de um motor de consultas e bases de conhecimento	Ontologia valida correlações mapeadas do esquema local para o global; uma sub-consulta esta correta se pertence a consulta global	Lógica descritiva (CARIN)	Ontologia representa uma cópia da estrutura (por ex. esquema de banco de dados) e a define melhor adicionando conceitos.
<i>DWQ</i>	Um modelo conceitual (em um nível superior) descreve as relações entre os modelos fonte e um modelo genérico.	Análise de conteúdo e descrição do processo de integração.	Lógica descritiva (extensão DLR)	Aceita abordagem por uma única ontologia global, por várias ontologias ou por uma abordagem híbrida.
<i>ONTOBROKER</i>	Meta-anotações que adicionam semântica às fontes de dados e evitam redundâncias; consiste de interface, motor de inferências ⁸⁴ e agente de coleta.	Representação de fontes de informação e armazenamento das meta-anotações.	Lógica descritiva (F-Logic)	Oferece um a possibilidade de criar uma linguagem de consulta e possui uma linguagem de meta-anotações.
<i>SHOE</i>	Meta-anotações que adicionam semântica às fontes de dados.	Expressam as meta-anotações com o uso direto ou a extensão de ontologias presentes na Internet.	Extensões ao HTML	As meta-anotações possibilitam ao autor do documento especificar metadados.
<i>BUSTER</i>	Regras de interação orientam a correlação entre as fontes em função de uma visão integrada produzida pelo mediador.	Ontologia global representa um vocabulário compartilhado que define faixas de atributos e conceitos.	Lógica descritiva (OIL)	Ontologia representa uma cópia da estrutura (por ex. esquema de banco de dados) e a define melhor adicionando conceitos.

Figura 10 – Quadro resumo de projetos que utilizam ontologias

⁸⁴

Um programa que infere novos fatos a partir de fatos conhecidos, utilizando-se de regras. Em geral faz parte de um interpretador *Prolog*, *sistema especialista* ou sistema baseado em conhecimento.

CAPÍTULO 5 – A PROPOSTA PARA UM META-MODELO DE INTEGRAÇÃO

5.1 Introdução

Ao se falar sobre construção de um modelo, pode-se estar tratando de uma ampla gama de possibilidades, já bastante discutidas na literatura. Entretanto, algumas características podem ser consideradas comuns aos modelos: mapeamento (são representações dos “originais”), redução (não se mapeiam todos os atributos do original) e pragmatismo (nem sempre pertencem a mesma classe dos originais e cumprem suas funções orientados para objetivos) (SAYÃO, 2001). Em função da interdisciplinaridade característica da Ciência da Informação, o modelo aqui proposto tem relações com modelos desenvolvidos em outros ramos da ciência.

A maioria dos trabalhos referentes à arquitetura de sistemas de informação incorpora modelos, implícita ou explicitamente. O modelo que se pretende nesse trabalho pode, além de suas funções tradicionais, auxiliar no gerenciamento de mudanças (nos esquemas das fontes e possibilidade de inserção de novas fontes), que é característica dos sistemas atuais de informação e proporcionar a separação entre funções de alto nível, mais próximas ao usuário (por exemplo, interfaces), das funções de baixo nível, ocultando possíveis problemas de heterogeneidade.

No contexto de aplicação do modelo, existem dois tipos de buscas possíveis: as federadas e o *harvesting* (literalmente “colheita”). Nas buscas *federadas*, os usuários submetem uma consulta ao sistema de busca e este a executa em vários locais (ambientes da Web), delegando a cada sistema a execução da consulta do usuário. Já nas buscas *harvesting* (BOWMAN et al., 1995), por algum meio (um protocolo, por exemplo), os dados ou metadados são trazidos das bases originais e armazenados em um repositório local, onde as buscas são efetuadas. O *harvesting* é o princípio utilizado pelo projeto SABiO, originando assim, um repositório com diversos tipos de tabela/bases. Essa situação acontece também em um ambiente organizacional, onde um *warehouse* ou *Enterprise Resource Planning (ERP)*, recebe dados de diversas bases localizadas dentro da empresa e devem integrá-las.

A partir da pesquisa em trabalhos relacionados, a Seção 4.3 apresentou diversos projetos que utilizam ontologias para integração. Uma dessas abordagens foi escolhida para nortear este processo de integração no meta-modelo em construção⁸⁵. O *DataWarehouse Quality*

⁸⁵

A combinação de abordagens com o objetivo de aproveitar as melhores características de projetos diferentes pode ser objeto de trabalho futuro ou complementar.

(DWQ) (CALVANESE et al., 1998) apresentado na Seção 4.3.9, se propõe a executar a tarefa de integrar as diversas bases de um *warehouse* através de uma ontologia de alto nível (que representa todas as bases e onde as consultas são feitas) e um processo de integração definido por lógica descritiva. Nesse trabalho, faz-se uma analogia entre o ambiente do SABiO e um *warehouse* organizacional, de forma que a integração no projeto SABiO possa ser tratada de forma semelhante.

Antes de descrever o meta-modelo, será necessário o entendimento de alguns conceitos essenciais para sua construção, efetivada a partir de diferentes técnicas:

- As relações entre as classes determinadas para o meta-modelo foram descritas em OIL; alguns conceitos básicos sobre o OIL são apresentados nos Anexo C;
- O meta-modelo será escrito em sua forma final utilizando-se *Resource Description Framework Schema (RDFS)*. Alguns conceitos básicos do RDF/RDFS também são apresentados no Anexo C;
- A abordagem de integração no nível conceitual baseia-se nos trabalhos de Calvanese et al. (1998) sobre integração em *data-warehouses*. O autor propõe uma extensão à lógica descritiva, a qual considera suficientemente expressiva para representar a integração entre os diferentes modelos que representam as fontes heterogêneas de dados. Considerações introdutórias sobre a lógica descritiva e sua terminologia básica são apresentadas no Anexo A;
- A descrição e o formalismo matemático que representam o modelo integrado é apresentado no Anexo B;
- As Seções 5.4.1 e 5.4.2 apresentam as diretrizes para construção do meta-modelo; sua sintaxe em OIL e RDFS é apresentada no Anexo D;
- O Anexo B apresenta um pequeno exemplo que ilustra como a lógica descritiva foi utilizada para a tarefa de integração; além disso discorre brevemente sobre a importância do tipo de representação do conhecimento utilizado para execução de deduções, objetivando obter melhor organização e gerenciamento da informação.

Caso haja interesse no formalismo matemático em lógica descritiva, deve-se ler em primeiro lugar o Anexo A – *Algumas definições e símbolos da lógica descritiva* (para os não iniciados no assunto) que traz as bases da sintaxe da lógica. Em seguida, o Anexo B – *O processo de integração baseado em lógica descritiva* pode ser lido por aqueles que desejam conhecer detalhes do processo de integração. Caso já se tenha algum conhecimento sobre OIL e RDFS pode-se dispensar a leitura do Anexo C, pois trata-se de uma descrição bastante elementar que tem o objetivo de dar informações básicas para aqueles que não conhecem o assunto.

O capítulo se estrutura da seguinte forma: na Seção 5.2 faz-se a distinção entre modelos e meta-modelos, definindo este último no escopo desse trabalho. A Seção 5.3 apresenta uma proposta preliminar de uma ontologia para o domínio de registros bibliográficos, através das várias tarefas que compõem a atividade. A Seção 5.4 apresenta informalmente o meta-modelo, sua arquitetura e representação. O formalismo, conforme citado, está nos Anexos.

5.2 Modelos e meta-modelos

Para esclarecer a diferença entre modelos e meta-modelos, citados até aqui quase que indistintamente, algumas considerações devem ser feitas. De modo geral, um meta-modelo apresenta as definições e os relacionamentos para um modelo, descrevendo-o. Em geral a meta-modelagem é baseada em quatro camadas (*METAOBJECT FACILITY*, 2002):

- A camada de objetos (do usuário), que consiste da informação que se deseja descrever (normalmente chamada de “dado”);
- A camada de modelo, que consiste de metadados que descrevem informações (são em geral agregados informalmente aos modelos);
- A camada de meta-modelo, que consiste das descrições (isto é, metadados) que definem a estrutura e a semântica dos metadados. Meta-meta-dado é informalmente agregado aos meta-modelos. Pode-se imaginar um meta-modelo como uma “linguagem” para descrever diferentes tipos de modelos;
- A camada meta-meta-modelo é composta de descrições da estrutura e semânticas de meta-meta-dados. Em outras palavras, é uma linguagem para definir diferentes tipos de metadados.

Kerhervé (1997)⁸⁶ citado por Barreto (1999), também apresenta uma arquitetura de meta-modelagem de quatro níveis, semelhante à apresentada anteriormente.

Pelo menos em teoria, é possível se ter “meta-meta-meta-modelos” e assim por diante. Isso não é necessário na maioria dos casos, mas abre a possibilidade de descrição da estrutura do conhecimento até o nível desejado.

Um meta-modelo é uma definição precisa das construções e regras necessárias para criar modelos semânticos (*METAMODEL*, 2002). Um meta-modelo é sempre construído para um propósito em particular e não é razoável tentar entendê-lo sem conhecer o objetivo para o qual foi criado.

⁸⁶ KERHERVÉ, B. *Models for Metadata or Metamodels for Data?* Disponível em: <<http://computer.org/conferen/proceed/meta97/papers/bkerherve/bkerherve.html>>. Acesso em: 20 set. 2002.

De forma a uniformizar a utilização do termo no contexto desse trabalho, adota-se a seguinte definição: *um meta-modelo é um modelo que descreve outros modelos a partir de regras necessárias para especificar sua semântica*. O objetivo do meta-modelo é descrever os componentes de um modelo e seus relacionamentos.

Este trabalho desenvolve um meta-modelo, o qual pretende lidar com o significado dos meta-dados disponíveis para um sistema e com o seu relacionamento, objetivando a interoperabilidade entre fontes distintas. Em particular, o propósito aqui é apresentar os conceitos e relacionamentos entre os itens mais comuns de referências bibliográficas, de forma a integrar as diversas fontes de dados consultadas pelo SABiO.

Na Seção 5.4 é apresentado o meta-modelo que faz a integração entre o modelo de dados do SABiO (aqui chamado de modelo-genérico) e os modelos arbitrários de outras fontes de informação (aqui chamados de modelos-fonte). O meta-modelo será apresentado em sua forma final na sintaxe RDF/RDFS. A Seção seguinte propõe uma ontologia que também faz parte da arquitetura de integração.

5.3 Uma proposta preliminar para uma ontologia bibliográfica

5.3.1 Introdução

O meta-modelo consiste de uma ontologia, de alto nível, onde são definidas as relações entre as classes de dados bibliográficos. Trata-se de uma ontologia genérica que tem por objetivo possibilitar o mapeamento entre os modelos das fontes, visando essencialmente a integração. Nesse caso, a ontologia serve apenas como ferramenta para se obter interoperabilidade. O meta-modelo será apresentado na Seção 5.4.

O papel de uma ontologia é possibilitar expressividade em relação ao domínio do conhecimento que pretende representar. Apresenta-se aqui uma proposta para uma ontologia bibliográfica que deve permitir responder perguntas importantes sobre o domínio. Essa ontologia também é parte da arquitetura do meta-modelo e participa do processo de integração, pois será utilizada como modelo genérico para o SABiO. A participação da ontologia bibliográfica no meta-modelo será melhor explicada na Seção 5.4.

O domínio da ontologia bibliográfica diz respeito à referências bibliográficas acadêmicas e seu objetivo é fornecer informações para a tomada de decisão. 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, muitas perguntas devem ser respondidas. Estudos

bibliométricos são necessários para que se possa estudar quais os indicadores seriam mais adequados.

A construção de uma ontologia tão sofisticada como esta, está além do escopo deste trabalho, podendo ser realizada em trabalho futuro. Não se tem a pretensão de desenvolver aqui uma ontologia que seja consenso ou padrão para referências bibliográficas. Trata-se de uma tarefa complexa, que envolve uma equipe inter-disciplinar, demanda prazo, recursos e conhecimento sobre o domínio, provavelmente provenientes de especialistas que devem ser alocados para a tarefa.

Com o objetivo de ilustrar a construção de uma possível ontologia bibliográfica e formular uma primeira proposta para o modelo genérico do SABiO, apresenta-se em seguida um roteiro. Entretanto, conforme citado, a ontologia construída representa mais uma ilustração de como o processo pode ser conduzido, e, obviamente, tem limitações, não podendo atender a uma ampla gama de necessidades.

Para a definição da ontologia bibliográfica, seguem-se algumas recomendações como: determinar o domínio e o escopo da ontologia, considerar o reuso de ontologias existentes, enumerar os termos importantes da ontologia, definir as classes e sua hierarquia, definir as propriedades das classes, definir as restrições sobre as propriedades e criar instâncias (NOY; GUINNESS, 2001). Essas recomendações são utilizadas, na seqüência, como um roteiro para a definição de uma ontologia sobre referências bibliográficas.

5.3.2 Domínio e o escopo da ontologia

A ontologia tem com objetivo relacionar os diversos tipos de dados que podem representar partes de uma referência bibliográfica. Deve cobrir, assim, o domínio representado por todos os dados (e metadados) que podem ser encontrados em uma referência bibliográfica.

A ontologia é usada para organizar os conceitos do domínio e deve ser capaz de 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 de entrevistas informais com profissionais e professores da ECI-UFMG⁸⁷, além da consulta a alguns trabalhos

⁸⁷ Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

bibliométricos (JACOBS, 2001; HAWKINS, 2001; GEISLER, 1999)⁸⁸.

Uma ontologia sobre referências bibliográficas deve possibilitar a resposta de questões do tipo apresentado abaixo:

- 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 determinado artigo nos últimos X anos ?
 - ❑ Quais são os periódicos mais relevantes para uma área?
 - ❑ Quais os tipos de publicações mais produzidas (ensaio, resultados de pesquisa, artigo científico, relatório de pesquisa, relatório técnico, artigo em jornal científico, capítulo de livro, anais de conferencia)?
 - ❑ Quais os assuntos foram mais publicados e em que anos?
- Sobre a área de conhecimento:
 - ❑ Quais são as palavras chave da área?
 - ❑ Que número médio de publicações é produzido na área por ano?
 - ❑ Qual é o local (universidade ou país) em que a área produz mais publicações?
 - ❑ Como é a produtividade científica da área em relação a outras áreas do conhecimento?

Conforme já citado, estas perguntas podem fazer parte de um estudo bibliométrico, que, entretanto, demanda um investimento além das possibilidades deste trabalho. Um exemplo de utilização e apresentação de regras de bibliometria para medir o trabalho dos autores de Ciência da Computação é encontrado no *Citeseer*⁸⁹, um índice de citação autônomo. Um dos recursos da ferramenta é um gráfico onde se mostra a relação do número de citações de um artigo versus o ano de publicação dos artigos citados.

Após determinar o escopo da ontologia que se pretende desenvolver, o próximo passo é consultar se já existem ontologias desenvolvidas no domínio de interesse. Como a construção de uma ontologia é uma tarefa que requer esforço, qualquer trabalho anterior pode ajudar.

⁸⁸ Nesse trabalho não se faz uma revisão bibliográfica sobre bibliometria. As consultas foram limitadas e apenas com o objetivo de se entender importância da bibliometria na definição de escopo de uma ontologia.

⁸⁹ Disponível para consulta na Internet em: <<http://citeseer.nj.nec.com/lawrence99indexing.html>>.

5.3.3 Ontologias existentes

Uma pesquisa nas bibliotecas de ontologias apresentadas na Seção 3.2.5 (além de outras fontes complementares) mostra que existem diversas ontologias relacionadas a dados bibliográficos. Além das encontradas nesses repositórios disponíveis na Internet, é importante considerar que o padrão MARC contém informações relevantes e tradicionais sobre registros bibliográficos. Outros padrões, por exemplo de metadados, como o *Dublin Core*, também são levados em consideração. Além disso, e, mais importante, é necessário considerar as necessidades do ambiente em que a ontologia pretende atuar.

Para a construção da ontologia bibliográfica são utilizados como exemplos, informações de seis ontologias disponíveis na Internet, relacionadas a registros bibliográficos:

- (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>> (21 campos mais utilizados do MARC).
- (6) <<http://orlando.drc.com/SemanticWeb/Topics/Ontology/Ontologies.htm>> (*Dublin Core*).

Essas ontologias foram selecionadas para que se possa verificar quais são os termos mais freqüentes no domínio e assim utilizá-los na ontologia bibliográfica. O Padrão MARC não tem o formato de uma ontologia, como os outros exemplos disponíveis, mas seus campos dizem respeito especificamente a dados bibliográficos. É difícil pensar nesse domínio sem pesquisar o MARC.

Na seqüência, é criada uma lista abrangente de termos relacionados ao assunto, sem a preocupação de sobrepor conceitos, 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 candidatos a termos em uma ontologia de referências bibliográficas pode conter diversas possibilidades e algumas repetições. Uma proposta em forma de tabela é apresentada no Anexo F, onde se encontram os campos que aparecem nas ontologias já existentes (apresentadas e numeradas de 1 a 6 dessa seção), agrupados em ordem alfabética. A tabela também mostra qual o papel do conceito de ontologia (classe, subclasse, propriedade, etc). A partir dessa lista e das necessidades levantadas no ambiente de trabalho, serão apresentados, na próxima Seção, os conceitos e as hierarquias da nova ontologia em construção.

5.3.4 Definição das classes, da hierarquia e das propriedades

Uma abordagem *top-down* para determinação das classes deve partir dos conceitos mais gerais até alcançar os mais especializados. Em uma abordagem *bottom-up*, o caminho seria o inverso. Uma combinação das duas formas é a melhor alternativa e, dessa forma, adotada aqui.

Independente da abordagem escolhida (*top down*, *bottom-up* ou uma combinação), 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 na ontologia e vão se tornar referências na hierarquia. Organizam-se as classes em uma taxonomia hierárquica perguntando 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; GUINNESS, 2001, p. 6). Obviamente, termos que aparecem na lista podem também ser descartados.

Para montagem da ontologia, utiliza-se a ferramenta OIEd, um editor de ontologias desenvolvido na *Manchester University*⁹⁰. A seguir a Figura 11 mostra uma tela do editor onde se determinam as classes (ver lado esquerdo da tela).

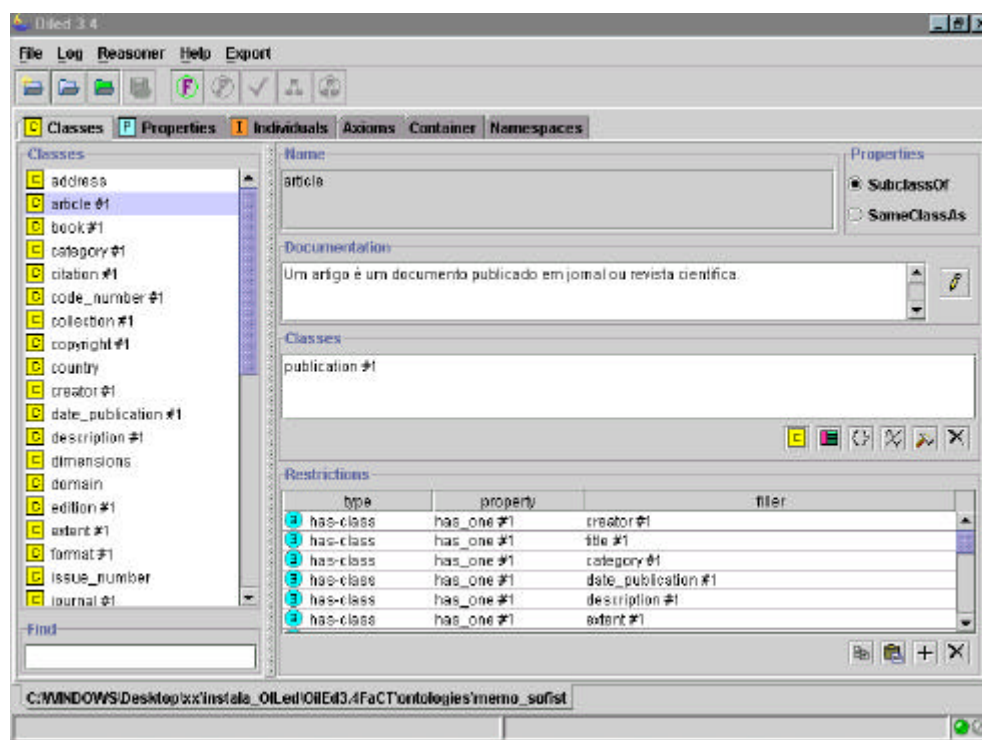


Figura 11 – Tela do editor de ontologias OIEd mostrando as classes para a ontologia bibliográfica

⁹⁰

Disponível na Internet em: <www.man.ac.uk>. Acesso em: 27 nov. 2002.

A partir da lista citada na seção anterior e apresentada na Figura D1 do Anexo D, as classes são criadas. O processo de utilizar termos da lista, definir as classes e subclasses é apresentado na Figura D2 do Anexo D, onde foi realizado um agrupamento semântico dos termos.

Além das classes que representam itens de acervo (*article*, *book*, etc.), outras classes que dizem respeito às características de um registro bibliográfico foram definidas no OILED, de forma similar à apresentada acima, mas não são apresentadas aqui para maior objetividade, pois não possuem subclasses⁹¹.

Entretanto, apenas classes não possibilitam respostas à perguntas efetuadas na Seção 5.3.2. Muitos dos termos que “sobram” na lista da Seção 5.3.3 após a definição das classes provavelmente são propriedades dessas classes. Estas propriedades são chamadas de *slots* e podem ser de diversos tipos: intrínsecas, extrínsecas, partes de um objeto estruturado e relações entre membros individuais de uma classe e outros itens (NOY; GUINNESS, 2001).

Todas as subclasses de uma classe herdam o *slot* daquela classe e este deve se referir sempre que possível, à classe mais geral que possui a propriedade. Na Figura 11 as propriedades da classe *article* estão definidas no canto inferior direito da tela. Os *slots* definidos para a ontologia bibliográfica são: *authoring*, *correcting*, *has-domain*, *has-one*, *is-a*, *is-published-by*, *publish-one*, *publishing*.

5.3.5 Definição das restrições sobre as propriedades

Os *slots* podem ter diferentes características, chamadas de *facets*, que descrevem o tipo dos valores, os valores permitidos, o número de valores (cardinalidade) e outras características. Alguns *facets* usuais são: cardinalidade do *slot*, o tipo de valor do *slot* (*slot-value type*), o domínio e a faixa (*range*) de valores de um *slot*.

A cardinalidade de um *slot* define quantos valores um *slot* deve ter. Na Figura 12 a classe *article* tem dois *slots* de cardinalidade um: *citation* e *copyright*. Isso indica que um artigo tem sempre, no mínimo, uma “citação” e um “direito autoral”.

⁹¹ Ver lista completa no Anexo F, Figura F3.

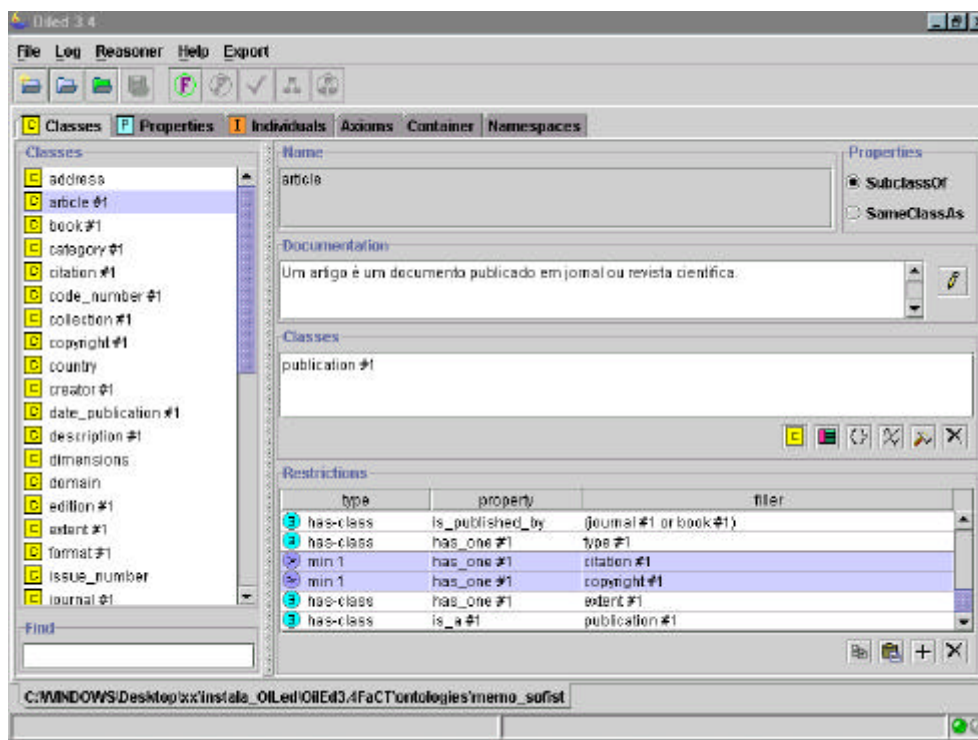


Figura 12 – Tela do editor de ontologias Oiled mostrando as restrições de cardinalidade um

O tipo de valor do *slot* (*slot-value type*) descreve que tipos de valores podem ser utilizados no *slot*. Os tipos de valores mais comuns são: *strings*, números, *slots* booleanos (representam simplesmente sim e não dentre duas possibilidades), *slots* enumerados (definem uma lista de valores permitidos para o *slot*) e do tipo instância (permitem definir relacionamentos entre indivíduos).

As classes permitidas para *slots* do tipo instância são em geral chamadas faixas (*range*) de um *slot*. As classes às quais um *slot* está anexado ou as classes que têm suas propriedades descritas para um *slot* são chamados de domínio (*domain*). A Figura F3 do Anexo F resume as classes e seus respectivos *slots* e *facets*.

A última atividade 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 dos *slots*.

5.3.6 Considerações finais

Conforme já citado, a construção de uma ontologia para um domínio do conhecimento é uma tarefa dispendiosa e está além das possibilidades desse trabalho. A ontologia bibliográfica foi construída com o objetivo de ilustrar o processo de construção e mostrar como poderia ser o formato de uma ontologia sobre o assunto. Para torná-la útil, ainda é necessário

pesquisa, a qual pode ser tema para um trabalho futuro. A ontologia apresentada é mais uma ilustração do que uma estrutura que representa o consenso na área.

Considerações sobre a definição das classes e hierarquias, sobre herança múltipla, sobre quando ou não introduzir uma nova classe, sobre a necessidade de criar uma nova classe ou uma propriedade, sobre instâncias versus classes, sobre a definição de propriedades, dentre outras, podem ser encontrados em Noy e Guinness (2001).

5.4 O meta-modelo para integração

5.4.1 A arquitetura de integração do meta-modelo

Existem disponíveis algumas ontologias relacionadas, conforme apresentado na Seção 5.3.3, mas nenhuma delas trabalha com classes que agregam a mesma informação de diversas fontes. Por exemplo, uma classe “*creator*” do meta-modelo deverá receber subclasses de autores, das diversas fontes envolvidas. Dessa forma, optou-se por construir outra ontologia mais genérica, dita de “alto nível”, conforme apresentado na sequência. O objetivo desta ontologia que, em última instância, representa o meta-modelo, é simplesmente servir como ferramenta para integração. Possui portanto, características distintas da ontologia desenvolvida na Seção anterior, cujo objetivo é expressar o conhecimento de um domínio.

Utiliza-se aqui uma abordagem declarativa⁹² para o problema da integração de dados cujo objetivo é modelar os dados nas fontes através de uma linguagem declarativa. Constrói-se uma representação unificada que é utilizada pelo sistema global para responder consultas.

Existem ainda dois fatores essenciais para o projeto e manutenção de aplicações de integração: a modelagem conceitual e a possibilidade de procedimentos de inferência sobre a representação conceitual (CALVANESE et al., 1998)⁹³.

O meta-modelo proposto nesse trabalho consiste de uma representação conceitual que objetiva eliminar ou reduzir conflitos semânticos, através da construção de uma ontologia comum, de mais alto nível, integrando as fontes de dados heterogêneas. O ambiente de testes para

⁹² A abordagem declarativa traz vantagens sobre a abordagem procedural, visto que possibilita a manutenção consistente e global das fontes de informação, as quais são componentes dos sistemas que devem ser integrados. A abordagem declarativa é utilizada, por exemplo, no *SIMS* (citado nesse trabalho). Na abordagem procedural, os dados são integrados de forma *ad-hoc*, em relação a um grupo pré-definido de necessidades de informação. As arquiteturas baseadas nessa proposta utilizam *wrappers* (que trabalham sobre fontes de dados, convertendo os objetos para um modelo comum) e mediadores (que obtêm informações de um ou mais *wrappers*, refinam, integram e resolvem conflitos entre dados provenientes de fontes diferentes). A abordagem procedural é utilizada, por exemplo, no *TSIMMIS* (citado nesse trabalho).

⁹³ Para informações sobre os procedimentos de inferência, ver Anexo B.

aplicação do meta-modelo, conforme citado, será o projeto SABiO.

Para efetuar a integração entre o modelo de dados do SABiO e das diversas fontes de dados bibliográficos, constrói-se uma arquitetura de integração composta por três camadas: a conceitual, a lógica (esquemas) e a física (dados). A solução para problemas de interoperabilidade no nível semântico, de interesse para esse trabalho, é estudada no nível conceitual. As camadas lógica e física não são abordadas aqui.

A camada conceitual (aqui nomeada de meta-modelo do domínio, ou, MMD) é constituída pelo *Modelo-Genérico*, daqui em diante chamado “MG” (que corresponde ao modelo do SABiO), e, um ou mais *Modelos-Fonte*, daqui em diante chamado “MFs” (para melhor entendimento ver Figura 11).

Depois de construído, o MMD corresponde a uma visão unificada dos MFs e do MG. Esta visão unificada contém a representação de todos os dados disponíveis nas diversas fontes, e, as consultas sobre ela proporcionam o “cruzamento” de dados de fontes diferentes e obtenção de novas consultas combinadas. Também podem ser aplicadas regras de inferência sobre o MMD, possibilitando deduções lógicas.

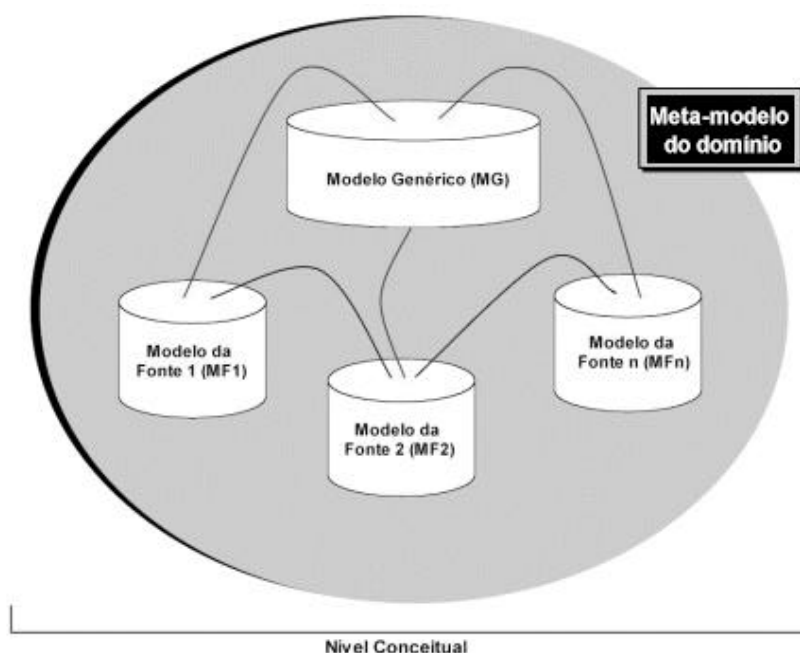


Figura 13 – Arquitetura da camada conceitual

O MMD consiste de uma representação dos conceitos e relações globais que são de interesse para o sistema. No caso em estudo, o interesse são dados bibliográficos de diversas fontes presentes na Internet. Cada MF corresponde uma fonte disponível, ou pelo menos parte de seus dados, visto que não é necessário que a uma fonte seja totalmente conceitualizada (a

abordagem é incremental).

O MMD é construído a partir do MG, após sua interação com os MFs. O MMD pode não ser uma representação completa de todas as fontes de dados (abordagem incremental), mas representa uma descrição consolidada dos conceitos e relações importantes até o momento. Tal descrição está sujeita a mudanças à medida que novas fontes de informação são integradas.

Além dos modelos acima, o MMD contém a especificação das interdependências entre elementos de diferentes MFs e, entre os MFs e o MG. O meta-modelo é o resultado final da integração, onde serão feitas as consultas do usuário. A Figura 14 apresenta os níveis de modelagem propostos:

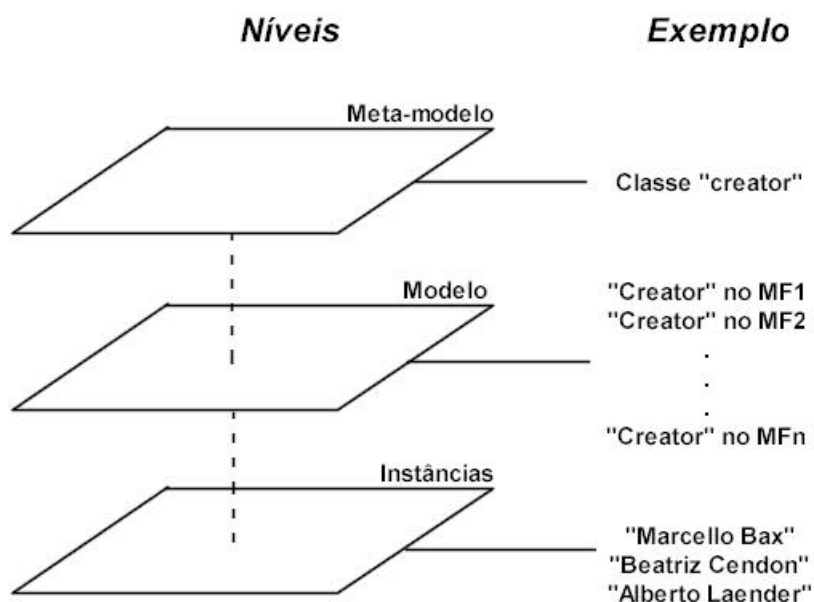


Figura 14 – Níveis de modelagem

O processo de integração deve mapear as relações entre os modelos (MF para MF e MFs para MG). Estas relações são expressas formalmente por meio da lógica descritiva, através de “*assertivas inter-modelos*” (CATARCI; LENZERINE, 1993) as quais possibilitam um mecanismo declarativo para representar dependências entre entidades, formalizando-as. A idéia é usar uma linguagem para expressar a noção de interdependências entre classes que pertencem a diferentes esquemas.

A formalização da noção de interdependência é obtida pelas *assertivas inter-modelo*, que são inter-relações conceituais entre os MFs e os MG. Tais interdependências permitem estabelecer relacionamentos entre a *definição intensional*⁹⁴ e classes representadas em esquemas

⁹⁴ Definição que expressa relacionamentos conceituais; o termo é explicado no Anexo A e no formalismo matemático apresentado no Anexo B.

diferentes. Por exemplo, pode-se afirmar, para uma dada linguagem, que o conceito representado pela classe *Estudante-Graduado* em um esquema S_1 é o mesmo que o conceito representado por *Estudante-Senior* em um esquema S_2 . Tal afirmativa implica em um tipo de equivalência *intensional* entre as duas classes; não implica que o grupo de instâncias do primeiro esquema seja sempre o mesmo que o grupo de instâncias do segundo. Ou seja, o “graduado” em uma escola equivale ao “sênior” em outra, mas os alunos não são necessariamente os mesmos.

O processo de integração envolve também a idéia de “*assertivas intra-modelo*”. Um modelo conceitual, como o definido anteriormente, é constituído por um grupo finito de *assertivas intra-modelo*, as quais expressam o conhecimento sobre as relações e conceitos presentes em cada modelo.

As condições em que as *assertivas inter-modelos* podem proporcionar integração são representadas formalmente em Lógica Descritiva. A forma em que se apresentam as *assertivas intra-modelo* e *inter-modelo* e detalhes sobre sua função no processo de integração, são explicadas e exemplificadas no Anexo B, que apresenta o formalismo matemático da integração. Uma breve descrição do raciocínio utilizado para se chegar à integração, em forma de roteiro, é apresentada na seqüência:

- 1) Um modelo conceitual, que corresponde ao meta-modelo do domínio (MMD), consiste de *assertivas intra-modelo*, que expressam o conhecimento dos *conceitos* e *relações* do modelo; e, de *assertivas inter-modelo*, que expressam o conhecimento do relacionamento entre fontes distintas;
- 2) As *assertivas intra-modelo* bem como as *relações* e os *conceitos* que elas descrevem, podem ser representadas formalmente em lógica descritiva;
- 3) Um conjunto de *assertivas intra-modelo* é satisfeita a partir da noção de *interpretação*⁹⁵ e de regras semânticas definidas;
- 4) No seu aspecto *intensional*⁹⁶, as *assertivas inter-modelo* devem atender a condições idênticas às *assertivas intra-modelo*; no aspecto *extensional*⁹⁷, a demonstração apresenta algumas diferenças;
- 5) Obtém-se o meta-modelo ao se conseguir que todas as *assertivas intra-modelo* e *assertivas inter-modelo* sejam atendidas (conforme passo 3);

⁹⁵ Consiste em um domínio e uma função que atribuem um valor da função a cada nome de conceito, nome de papel e nome de indivíduo; maiores detalhes vide Anexo A.

⁹⁶ Relativas a parte terminológica da base de conhecimento (T-Box); maiores detalhes sobre esse tipo de definição são encontrados no Anexo B;

⁹⁷ Relativas a parte assertiva da base de conhecimento (A-Box); maiores detalhes sobre esse tipo de definição são encontrados no Anexo B;

- 6) Utilizando serviços de inferência⁹⁸ sobre o meta-modelo, podem-se deduzir automaticamente consequências lógicas, mesmo que nem todas as relações envolvidas estejam presentes em todos os modelos.

5.4.2 Descrição informal do meta-modelo

Utilizando-se as considerações sobre a definição do meta-modelo feitas na seção anterior, representa-se o modelo inicialmente em OIL e em seguida em RDFS. Para a descrição da sintaxe em OIL utiliza-se novamente a ferramenta OILed. O meta-modelo é apresentado em seguida informalmente, da maneira em que foi introduzido no editor de ontologias. O próprio editor gera a sintaxe OIL e a RDFS, as quais são apresentadas no Anexo D. A Figura 15 apresenta a tela inicial do OILed, com as classes do meta-modelo já definidas:

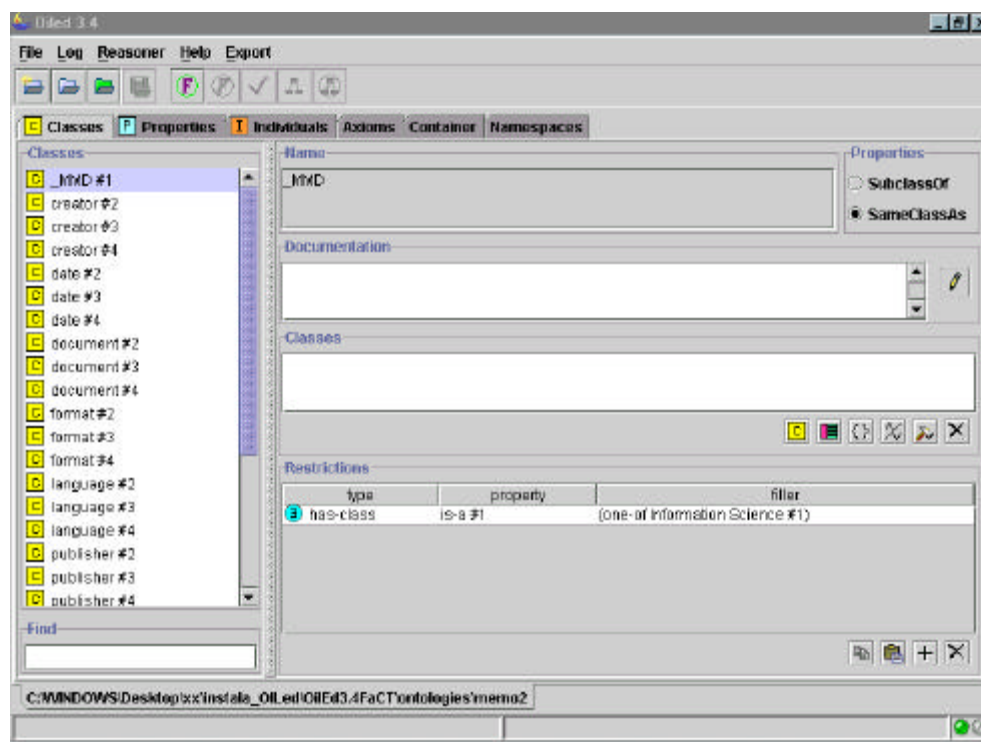


Figura 15 – Tela do OILed, editor de ontologias utilizado para construção do meta-modelo

Apresenta-se em seguida a descrição informal do meta-modelo, para que se possa compreender sua concepção, antes que seja escrito na sintaxe definitiva (RDFS). A descrição informal ordena as entidades na mesma sequência em que são apresentadas pelo editor. Os sinais “#” que aparecem ao lado do nome da classe são colocados automaticamente pelo OILed e representam diferentes espaços de nomes (*namespaces*), ou seja, diferentes *URIs-Uniform Resource Identifier* para um recurso. Os *namespaces* estão definidos como #1 para a raiz

⁹⁸

Os serviços de inferência são brevemente discutidos ao final do Anexo C.

<www.eci.ufmg.br/sabio> (meta-modelo), #2 para <www.eci.ufmg.br/sabio/MG> (modelo-genérico), #3 para <www.eci.ufmg.br/MFn> (último modelo fonte considerado em uma integração) e #4 para <www.eci.ufmg.br/MF1> (primeiro modelo fonte considerado em uma integração). Com a descrição informal pode-se compreender as relações entre as classes do meta-modelo, as quais são formalizadas em seções posteriores. As classes foram criadas de forma análoga aos metadados do padrão *Dublin Core*⁹⁹, mas se encontram em um nível superior (meta-meta)¹⁰⁰.

▪ Definição das classes

<i>Nome da classe</i>	: _MMD
<i>Subclasses de</i>	: não tem (NT)
<i>Restrições</i>	: não tem (NT)
<i>Descrição</i>	: classe que especifica as fontes integradas.
<i>Nome da classe</i>	: creator #2
<i>Subclasses de</i>	: não tem (NT)
<i>Restrições</i>	: is-a person
<i>Descrição</i>	: classe que agrupa os autores dos recursos pesquisados, representando o modelo-genérico;
<i>Nome da classe</i>	: creator #3
<i>Subclasses de</i>	: creator #2
<i>Restrições</i>	: as mesmas da superclasse
<i>Descrição</i>	: classe que agrupa os autores do recursos pesquisados, representando o último modelo fonte pesquisado.
<i>Nome da classe</i>	: creator #4
<i>Subclasses de</i>	: creator #2
<i>Restrições</i>	: as mesmas da superclasse
<i>Descrição</i>	: classe que agrupa os autores do recursos pesquisados, representando o primeiro modelo fonte pesquisado.
<i>Nome da classe</i>	: date #2
<i>Subclasses de</i>	: NT
<i>Restrições</i>	: NT
<i>Descrição</i>	: data da referencia bibliográfica pesquisada, representado no modelo-genérico.
<i>Nome da classe</i>	: date #3
<i>Subclasses de</i>	: date#2
<i>Restrições</i>	: NT
<i>Descrição</i>	: data da referencia bibliográfica pesquisada, representado no último modelo-fonte.
<i>Nome da classe</i>	: date #4
<i>Subclasses de</i>	: date#2
<i>Restrições</i>	: NT
<i>Descrição</i>	: data da referencia bibliográfica pesquisada, representado no primeiro modelo-fonte.
<i>Nome da classe</i>	: document #2
<i>Subclasses de</i>	: NT
<i>Restrições</i>	: has-one title, creator, subject, publisher, date, format, language, creator; has-class is-published creator, publisher; : has-domain <i>Information Science</i>
<i>Descrição</i>	: classe que agrupa os documentos que representam as referências bibliográficas pesquisadas no sistema SABiO; definida no modelo-genérico;
<i>Nome da classe</i>	: document #3

⁹⁹ Padrão de meta-dados para uso na Internet. Disponível em: <<http://dublincore.org/>>.

¹⁰⁰ Os termos são definidos em inglês acompanhando a apresentação do *Dublin Core*.

Subclasses de : document #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os documentos que representam as referência bibliográficas pesquisadas no sistema SABiO; definida no último modelo-fonte pesquisado;

Nome da classe : document #4
Subclasses de : document #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os documentos que representam as referência bibliográficas pesquisadas no sistema SABiO; definida no primeiro modelo-fonte pesquisado;

Nome da classe : format #2
Subclasses de : NT
Restrições : has-one type#2
Descrição : classe que agrupa os formatos em que estão as referências bibliográficas pesquisadas no sistema SABiO; definida no modelo-genérico;

Nome da classe : format #3
Subclasses de : format #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os documentos que representam as referência bibliográficas pesquisadas no sistema SABiO; definida no último modelo-fonte pesquisado;

Nome da classe : format #4
Subclasses de : format #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os documentos que representam as referência bibliográficas pesquisadas no sistema SABiO; definida no primeiro modelo-fonte pesquisado;

Nome da classe : language #2
Subclasses de : NT
Restrições : is-a portuguese
Descrição : classe que define a linguagem dos documentos pesquisados como referência bibliográficas; definida no modelo-genérico pesquisado;

Nome da classe : language #3
Subclasses de : language #2
Restrições : mesmas da superclasse
Descrição : classe que define a linguagem dos documentos pesquisados como referência bibliográficas; definida no primeiro modelo-fonte pesquisado;

Nome da classe : language #4
Subclasses de : language #2
Restrições : is-a portuguese
Descrição : classe que define a linguagem dos documentos pesquisados como referência bibliográficas; definida no último modelo-fonte pesquisado;

Nome da classe : publisher #2
Subclasses de : NT
Restrições : publish-one document, has-class is-a company or person
Descrição : classe que agrupa os editores dos documentos pesquisados como referência bibliográfica; definida no modelo-genérico pesquisado;

Nome da classe : publisher #3
Subclasses de : publisher #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os editores dos documentos pesquisados como referência bibliográfica; definida no último modelo-fonte pesquisado;

Nome da classe : publisher #4
Subclasses de : publisher #2
Restrições : mesmas da superclasse
Descrição : classe que agrupa os editores dos documentos pesquisados como referência bibliográfica; definida no primeiro modelo-fonte pesquisado;

Nome da classe : subject #2
Subclasses de : NT

Restrições : NT
Descrição : classe que agrupa os assuntos dos documentos pesquisados como referência bibliográfica; definida no modelo-genérico pesquisado;

Nome da classe : subject #3
Subclasses de : subject #2
Restrições : NT
Descrição : classe que agrupa os assuntos do documento pesquisado como referência bibliográfica; definida no último modelo-fonte pesquisado;

Nome da classe : subject #4
Subclasses de : subject #2
Restrições : NT
Descrição : classe que agrupa os assuntos dos documentos pesquisados como referência bibliográfica; definida no primeiro modelo-fonte pesquisado;

Nome da classe : title #2
Subclasses de : NT
Restrições : NT
Descrição : classe que agrupa os títulos dos documentos pesquisados como referência bibliográfica; definida no modelo-genérico pesquisado;

Nome da classe : title #3
Subclasses de : title #2
Restrições : NT
Descrição : classe que agrupa os títulos dos documentos pesquisados como referência bibliográfica; definida no último modelo-fonte pesquisado;

Nome da classe : title #4
Subclasses de : title #2
Restrições : NT
Descrição : classe que agrupa os títulos dos documentos pesquisados como referência bibliográfica; definida no primeiro modelo-fonte pesquisado;

Nome da classe : type #2
Subclasses de : NT
Restrições : NT
Descrição : classe que agrupa os tipos dos documentos pesquisados como referência bibliográfica; definida no modelo-genérico pesquisado;

Nome da classe : type #3
Subclasses de : type#2
Restrições : NT
Descrição : classe que agrupa os tipos dos documentos pesquisados como referência bibliográfica; definida no último modelo-fonte pesquisado;

Nome da classe : type #4
Subclasses de : type#2
Restrições : NT
Descrição : classe que agrupa os tipos dos documentos pesquisados como referência bibliográfica; definida no primeiro modelo-fonte pesquisado;

▪ Definição das propriedades

Nome da propriedade : has-one
Subpropriedades de : NT
Descrição : associa um classe a outra conferindo idéia de posse

Nome da propriedade : is-a
Subpropriedades de : NT
Descrição : associa um classe a outra conferindo caracterização

Nome da propriedade : is-published
Subpropriedades de : NT
Descrição : associa uma classe *document* a uma outra classe *publisher* e/ou *creator*

Nome da propriedade : publish-one
Subpropriedades de : NT

Descrição : associa uma classe *publisher* e/ou *creator* à uma classe *document* (inverso de *is-published*).

Nome da propriedade : publishing

Subpropriedades de : NT

Descrição : associa uma classe *creator* a uma classe *document* (considerando-se que o autor pode publicar na Internet)

Nome da propriedade : authoring

Subpropriedades de : NT

Descrição : associa uma classe *creator* a uma classe *document*, expressando responsabilidade pelo conteúdo intelectual.

Nome da propriedade : correcting

Subpropriedades de : authoring

Descrição : associa uma classe *creator* a uma classe *document*, expressando responsabilidade pela revisão do documento.

Nome da propriedade : has-domain

Subpropriedades de : NT

Descrição : associa uma classe *document* ao domínio em estudo (*Information Science*).

▪ **Definição de atributos**

Nome do atributo: company

Descrição: empresa ou pessoa jurídica organizada conforme a legislação vigente.

Nome do atributo: Information Science

Descrição: área do conhecimento ou domínio do conhecimento pesquisado.

Nome do atributo: person

Descrição: pessoa, ser humano

Nome do atributo: portuguese

Descrição: linguagem natural que será considerada para pesquisa.

▪ **Axiomas**

Disjoint (document #2 | creator #2)
 Disjoint (document #2 | date #2)
 Disjoint (document #2 | language #2)
 Disjoint (document #2 | publisher #2)
 Disjoint (document #2 | title #2)
 Disjoint (document #2 | type #2)
 Disjoint (creator #2 | date #2)
 Disjoint (creator #2 | format #2)
 Disjoint (creator #2 | language #2)
 Disjoint (creator #2 | publisher #2)
 Disjoint (creator #2 | title #2)
 Disjoint (creator #2 | type #2)
 Disjoint (date #2 | format #2)
 Disjoint (date #2 | language #2)
 Disjoint (date #2 | publisher #2)
 Disjoint (date #2 | title #2)
 Disjoint (date #2 | type #2)
 Disjoint (format #2 | language #2)
 Disjoint (format #2 | publisher #2)
 Disjoint (format #2 | title #2)
 Disjoint (format #2 | type #2)
 Disjoint (language #2 | publisher #2)
 Disjoint (language #2 | title #2)
 Disjoint (language #2 | type #2)
 Disjoint (language #2 | title #2)
 Disjoint (title #2 | type #2)

▪ **Informações gerais (container)**

Title: Meta-modelo para integração.

Creator: Mauricio Almeida

Subject: meta-modelo para uso em projetos de integração de fontes de dados

Description: trabalho para dissertação de mestrado ECI – UFMG – 2002

Date: novembro / 2002

Versão: 1.0

- **Definições de namespaces:**

Default: endereço local de construção da ontologia

Namespace 1: <http://www.eci.ufmg.br/sabio#>

Namespace 2: <http://www.eci.ufmg.br/sabio/MG#>

Namespace 3: <http://www.eci.ufmg.br/sabio/MFn#>

Namespace 3: <http://www.eci.ufmg.br/sabio/MF1#>

A Figura 16 apresenta um fragmento das hierarquias definidas no metamodelo. O OILED permite apenas a apresentação de hierarquias de ontologias menores¹⁰¹, como é o caso.

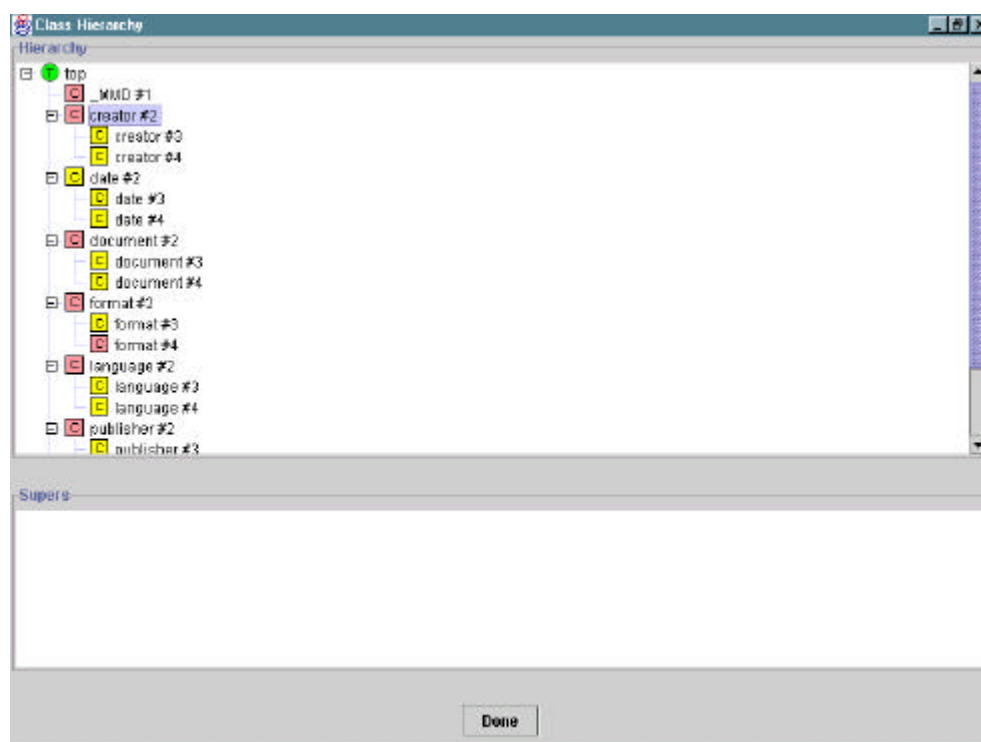


Figura 16 – Tela do OILED, com representação das hierarquias definidas entre as classes

5.4.3 Representação do meta-modelo

O meta-modelo proposto será construído inicialmente em OIL, onde são especificadas as relações e conceitos. Em seguida o OIL é traduzido em *RDFS*, utilizado no projeto SABIO.

A sintaxe RDF / XML não apresenta a expressividade requerida em algumas situações (conforme visto nas Figuras 5 e 6, Seção 2.2). O motivo para uso do *RDFS* está na necessidade de representar conhecimento adicional sobre o recurso¹⁰² (ou seja, no uso de

¹⁰¹ O manual de instalação software informa que o OILED não é um ambiente de desenvolvimento de ontologias completo e não suporta grandes ontologias.

¹⁰² A palavra “recurso” está sendo utilizada no contexto de acordo com seu significado em RDF.

metadados mais expressivos). O *RDFS* torna possível verificar se as propriedades aplicadas a um recurso fazem sentido e se os valores utilizados em conjunto com as propriedades são possíveis. Pode-se verificar a integridade dos dados, a validade de declarações, criar uma hierarquia de tipo definindo-se classes e subclasses, utilizar recursos de herança (comuns nas linguagens orientadas à objeto).

Um esquema pode ser especificado na sintaxe RDF, e partir daí, transportado e armazenado em outra forma e em outra sintaxe, pois o que importa nesse caso é o modelo que suporta a sintaxe e não a sintaxe em si. O meta-modelo neste trabalho é construído em *RDFS*, não significando que essa sintaxe seja obrigatória.

O RDF permite a apresentação textual e a apresentação gráfica, que mostra mais claramente os recursos e propriedades do meta-modelo, bem como suas inter-relações. Na sequência o RDF do meta-modelo será apresentado na notação gráfica. As marcações RDFS presentes baseiam-se em Nebel (1998), que possibilita algumas extensões.

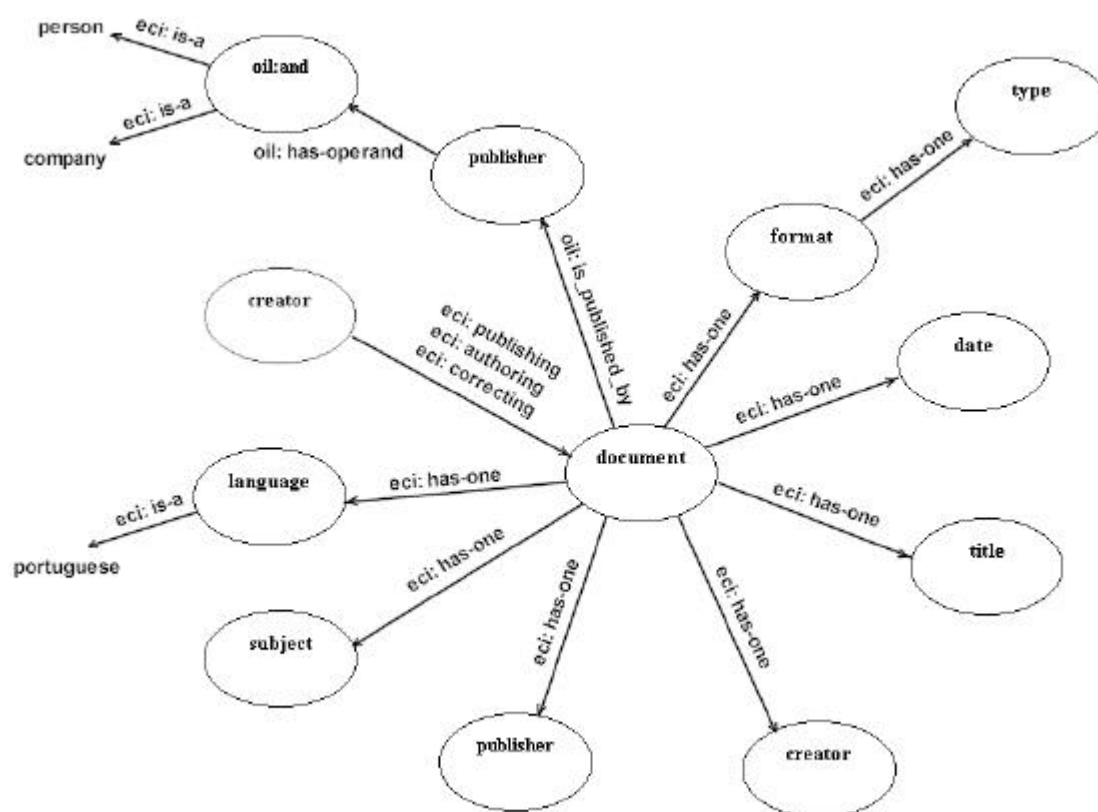


Figura 17 – Representação gráfica em RDF

A descrição final do meta-modelo, correspondente ao código RDFS que o representa e possibilita sua implementação, encontra-se no Anexo E.

CAPÍTULO 6 – DISCUSSÃO

Uma discussão sobre assunto tão amplo pode incluir considerações variadas. A interoperabilidade é hoje um assunto associado a questões importantes no âmbito informacional das organizações e instituições. Como se pode verificar nos capítulos iniciais deste trabalho, o assunto envolve questões que nem sempre são fáceis de tratar. A idéia de interoperabilidade humana, da relação com a pragmática, está além do escopo desse trabalho e poderia ser um assunto de pesquisa para abordagens cognitiva e social da informação.

Dessa forma, uma providência salutar no início do trabalho foi, o mais breve possível, delimitar escopo do problema ao nível semântico. A proposta foi estudar a integração semântica pois acredita-se que a integração sintática dificilmente ocorrerá, uma vez que as tendências apontam para ambientes cada vez mais heterogêneos nos níveis dos dados e da informação. Ao abordar o significado, espera-se que os “símbolos” já sejam compreendidos pelas entidades que se comunicam.

A dificuldade em acessar fontes de dados em função de sua heterogeneidade é um fato comum atualmente em ambientes de rede que caracterizam a troca de dados entre organizações. Existem, em qualquer local de trabalho, inúmeras diferenças que impedem a comunicação adequada. Pode-se imaginar uma analogia entre a heterogeneidade de fontes de dados e as dificuldades de comunicação entre as pessoas: cada pessoa é única, com uma cultura diferente, uma história peculiar, valores enraizados e experiências que tornam aspectos mais simples da troca de informações um complicado processo de entender “o que é o outro”. Diversos autores tratam desse assunto¹⁰³.

6.2 O contexto de estudo do nível semântico

Para que sistemas digitais possam tratar de questões semânticas, contribuições da Inteligência Artificial devem ser consideradas. A Ciência da Informação participa em estudos dessa natureza através das técnicas de organização do conhecimento, como por exemplo, as ontologias, taxonomias, *thesaurus* e outras estruturas. Por sua característica multidisciplinar, a Ciência da Informação se utiliza considerações de outras áreas, e, pode dar uma contribuição relevante na organização das bases de dados para uma recuperação mais eficiente. O objetivo é conferir sentido a entidades presentes nos sistemas digitais e bases de dados, visando em última instância, o usuário.

¹⁰³ Um exemplo são os trabalhos de Pierre Levy.

Não é possível se restringir a uma abordagem estritamente tecnológica, mas o conhecimento da tecnologia é inerente ao estudo dos sistemas digitais aqui tratados. Além disso não é mais razoável tratar dados por quaisquer técnicas, sem utilizar as ferramentas da tecnologia. Os ambientes de dados atuais são diferentes, maiores, mais complexos, e mais heterogêneos do que aqueles que os primeiros profissionais de organização da informação encontraram. Obviamente, por exemplo, uma indexação humana é de qualidade muito superior à outra executada pela máquina. Mas ambientes como a Internet, praticamente obrigam a adoção de sistemas automatizados para essa tarefa, pela inviabilidade de um tratamento manual.

A busca de inter-operabilidade atualmente está relacionada a um ambiente no qual a competitividade das empresas e das nações depende de sua capacidade de utilizar as informações disponíveis. Para tal, é necessário em primeiro lugar acessar os dados. Depois, tentar reduzir a sobrecarga de informações sobre as pessoas, que ocorre em função do crescimento da capacidade de disseminação de dados. Para reduzir esta sobrecarga, é preciso construir sistemas “mais racionais” que possam executar, pelo menos em parte, a tarefa de conferir sentido aos dados.

Assim, se procura combinar dados de forma que as pessoas possam gerar informações de forma mais eficientes, sem o desgaste proveniente de uma sobrecarga. Como exemplo, pode-se indagar se uma pessoa tem capacidade de absorver os resultados obtidos após uma consulta a mecanismos de busca da Internet. Entretanto, uma busca desse tipo seria difícil e trabalhosa caso tais mecanismos não existissem. Estudar possibilidades de que computadores e sistemas compreendam significados não é conferir-lhes mais importância, mas utilizá-los de forma mais eficiente como ferramenta.

6.3 A solução adotada

A necessidade de inte-operar e a possibilidade de abordar o problema pelo uso das ontologias foi um aspecto importante citado ao longo de todo o trabalho, cujo objetivo era chamar a atenção para a importância do assunto. Mesmo com o desenvolvimento constante das técnicas computacionais, acredita-se que deva existir uma limitação no desenvolvimento de algoritmos cada vez mais rápidos. Parece razoável, dessa forma, estudar as outras partes envolvidas no processo de fornecer dados aos usuários: a organização das bases de dados (a qual se aborda indiretamente nesse trabalho) e os problemas humanos (bem mais complexos e nem sempre passíveis de solução).

A diversidade de projetos de integração que utilizam ontologias demonstra a atenção que o assunto tem recebido da comunidade científica. O desenvolvimento de ferramentas como o

OILED, o *Protege*¹⁰⁴, dentre outros, utilizados na construção de ontologias, também demonstra o interesse em facilitar o acesso às técnicas pelos profissionais de informação. Além disso, essas ferramentas podem gerar a ontologia em códigos que são referências internacionais, como o OIL e o RDFS. Tal funcionalidade foi utilizada no meta-modelo aqui desenvolvido. Cabe citar também a facilidade de tradução entre a lógica descritiva e a linguagem OIL (com a qual se constrói a ontologia) e os padrões RDF/RDFS.

A idéia de combinar lógica, ontologias e computação para busca de melhores resultados de inter-operabilidade parece bastante aceitável. Pôde-se observar que a associação das três áreas é benéfica em uma abordagem de inter-operabilidade.

Um aspecto prático que vale ressaltar é o tipo de ontologia utilizada para integração. A solução apresentada utiliza uma ontologia de alto nível. O meta-modelo corresponde a uma ontologia global que mapeia todas as classes e relações das outras ontologias que compõem o ambiente, ou seja, o modelo do próprio SABiO e os modelos das fontes heterogêneas. Uma vantagem na integração da forma que é executada pelo meta-modelo é que o número de bases pode variar, visto que se trata de uma abordagem incremental. Por isso o sistema considera modelos-fonte de 1 até n.¹⁰⁵

Outras abordagens ao uso de ontologias para integração são possíveis, conforme mostram os projetos citados na Seção 4.3. Considera-se que construir uma ontologia de alto nível, mantendo atrelados ao seu formalismo os modelos das fontes de dados, é uma boa solução, pois limita as interferências de usuários na definição da semântica. Não é necessária a intervenção do usuário na construção da ontologia global. Caso o usuário possa definir o mapeamento, pode-se não ter controle sobre a semântica original e ela pode ser alterada. Outro aspecto a ser considerado é que a ontologia deve atuar tanto no mapeamento entre fontes de dados, quanto entre as fontes de dados e a aplicação.

Parece simples manter a ontologia de alto nível mapeando todas as possibilidades para uma eventual consulta e garantir que todas as outras ontologias (das fontes) atendam ao formalismo que representa a integração. Isso pode ser feito utilizando-se a idéia de herança dos

¹⁰⁴ O *Protege* é um outro editor de ontologias.

¹⁰⁵ No caso do projeto SABiO, utilizado com ambiente de testes, as possíveis fontes a considerar, visto que se trata de pesquisa bibliográfica, em Ciência da Informação e em português, são: Bireme. (<www.bireme.br>), na área de ciência e tecnologia, o Scielo (a base de conteúdo integral); Centro de Informações Nucleares (CIN/CNEN) (<<http://www.cnem.gov.br>>), na área de ciência e tecnologia, o Catálogo Coletivo de Anais de Congresso; Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) (<www.ibict.br>), bases bibliográficas do Catálogo Coletivo Nacional de Publicações Seriadas, Sistema de Teses e Divulgação Científica, Prossiga (contém bases de dados e links).

conceitos da ontologia de alto nível, resolvendo conflitos e ambigüidades. Ao mesmo tempo em que se permite estabelecer conexões entre conceitos de ontologias diferentes em termos de superclasses, não se estabelece uma correspondência direta entre os termos, podendo ocorrer algum tipo de perda de informação.

Para a computação requerida pela solução, existem indicações de que algoritmos e técnicas de processamento estão bem desenvolvidas para a lógica descritiva utilizada no processo de integração. Também para o execução de inferências, a lógica descritiva parece ser uma boa solução. Uma das características de destaque é a existência de análises de complexidade computacional detalhadas, relativas a algoritmos de inferência e problemas lógicos. Para essas análises a maioria das lógicas descritivas têm algoritmos de inferência ótimos e exemplos práticos implementados em vários sistemas atuais.

6.4 Questões pendentes e trabalhos complementares

Uma possibilidade que se apresenta interessante e que poderá ser avaliada é a adoção de abordagens conjuntas. Pode-se combinar a criação de uma ontologia de alto nível a outra que leve em conta aspectos lingüísticos para a execução do mapeamento. O *OBSERVER*, apresentado na Seção 4.3.5, adota essa última abordagem, levando em conta sinônimos, homônimos, etc. O sistema poderia, por exemplo, inferir que um *estudante-graduado* em uma fonte de dados tem o mesmo significado que *estudante-senior* em outra. Acredita-se que complementar à solução adotada com outras possibilidade apresentadas em projetos da área pode ser significativo para os objetivos de buscar inter-operabilidade. Além disso, outra abordagem que parece promissora é a que executa marcações nos documentos. O *ONTOBROKER* (seção 4.3.10) e o *SHOE* (seção 4.3.11) utilizam meta-anotações que adicionam semântica às fontes de dados.

Outra questão não verificada diz respeito a que tipo de computação pode ser necessária para implementar a solução. Conforme se pode observar pelas regras semânticas exigidas para integração (ver Anexo B) as operações estão relacionadas a *matchings* e montagem de um arquivo ou repositório que contenha as relações e conceitos de todas a ontologias envolvidas. Não se sabe sobre a viabilidade dessa computação se acontecer um aumento significativo do número de fontes que devem ser integradas.

A possibilidade de execução de inferências¹⁰⁶ torna o assunto abordado atraente. As inferências, são deduções que o sistema poderá fazer, a partir de fatos conhecidos e devidamente

¹⁰⁶ Os serviços de inferência são brevemente abordados após o formalismo que representa a integração, no Anexo B.

relacionados. Algumas inferências são necessárias para comprovar se características do modelo atendem às exigências. Por exemplo, no caso da lógica descritiva, alguns autores acreditam que os serviços de inferência típicos necessários no nível conceitual para que se possa efetuar integração, podem ser reduzidos à verificação da satisfabilidade do meta-modelo. Entretanto, capturar propriedades significativas no meta-modelo pode tornar a inferência uma tarefa complexa.

Serviços de inferência também são de extrema importância para manipulação de bases do conhecimento. Deve-se estudar o que o sistema poderá “inferir” por conta própria a partir de fatos estabelecidos. Em um domínio que tem como fatos as relações entre referências bibliográficas, pode-se obter deduções significativas. Um exemplo seria a situação de avaliação de normas da CAPES para programas de pós-graduação. Todo programa de ensino que deseja manter nota 5 deve considerar que todos os professores devem publicar no mínimo dois artigos científicos por ano. O sistema pode inferir que se um professor não publicou dois artigos no período, a nota da avaliação não poderá ser 5. Essa é a idéia da abordagem declarativa, em oposição à procedural, na qual seria necessário prever todas as possibilidades.

6.5 Propostas para aplicação do presente trabalho

O meta-modelo é utilizado no projeto SABiO, mas já se imaginam também outras aplicações. Quaisquer ambientes que sejam compostos por diversas fontes heterogêneas podem utilizar a idéia do meta-modelo como solução para integração. Imaginando-se abordagens conjuntas como as citadas na seção anterior, os resultados poderão ser mais significativos. Como idéia de aplicação prática da solução apresentada aqui, imagina-se uma possível participação em projetos de gestão do conhecimento. Como esses projetos visam a uma maior disseminação e acesso dos usuários a informações de toda a empresa, pode-se inferir a importância de técnicas que possibilitem a integração de dados de diversos setores e sua apresentação aos usuários finais em uma visão única e integrada.

Outra possibilidade é pensar na idéia de integração entre fontes de dados aplicada à integração inter-ontologias. O objetivo seria, dentro de um mesmo domínio do conhecimento, uniformizar os conceitos semanticamente iguais, independente das peculiaridades de cada ontologia individual. Esse processo não seria feito manualmente, de comum acordo com os construtores de cada ontologia, mas automaticamente à medida das necessidades em cada domínio do conhecimento.

Além dos trabalhos complementares citados, que devem abordar questões ainda pendentes relacionadas a inter-operabilidade de fontes heterogêneas, existem ainda idéias para trabalhos futuros, mais abrangentes, que tomariam como ponto de partida o que foi desenvolvido

aqui. Essa possibilidade de trabalhos futuros é apresentada na Seção 7.

Em última instância, as considerações sobre fontes de dados, lógica, ontologias e inteligência artificial, inseridas no âmbito da Ciência da Informação, pretendem possibilitar a projetos de inter-operabilidade, a capacidade de incrementar estratégias de gestão das informações e conhecimento. Para gerir conhecimento, é preciso antes adquirir, depois organizar, em seguida entender o sentido, integrar, e, finalmente, colaborar.

CAPITULO 7 – CONCLUSÕES

A idéia inicial desse trabalho era possibilitar a integração de fontes de dados heterogêneas bibliográficas conforme proposto na introdução. Iniciou-se a pesquisa com uma revisão d literatura objetivando entender de que trata a inter-operabilidade. Essa revisão foi ampla em função da variedade de abordagens que estudam o assunto. Foram mostrados diversos trabalhos relacionados que utilizam ontologias como ferramenta para integração. Escolheu-se uma linha dentre as abordagens pesquisadas (representação por lógica descritiva) e utilizou-se de sua idéia. Construi-se uma ontologia de alto nível (meta-modelo) com o mapeamento das fontes envolvidas. Em seguida, usou-se o meta-modelo no ambiente do SABiO para demonstrar a viabilidade de aplicação da solução. Apresentaram-se também exemplos, informações e atividades relacionadas à construção de ontologias, além de projetos passíveis de compor abordagens conjuntas. A seqüência dessas atividades explicita o raciocínio que conduziu a execução deste trabalho.

Foram estudadas as principais características de fontes de dados e a necessidade de inter-operação em ambientes heterogêneos, além das ontologias e suas características O estudo desses tópicos proporcionou uma visão do estado da arte na área.

A hipótese citada ao longo de todo o trabalho de que as ontologias podem ser úteis em projetos de integração, parece se confirmar sem dúvida ao final. Isso já parecia claro durante o estudo, até mesmo em função da quantidade de projetos e linhas de pesquisa que buscam integrar por intermédio dessa solução. Acredita-se que os resultados obtidos sejam razoáveis e utilizar ontologias pode ser uma opção para lidar com a complexidade dos ambientes atuais e prover inter-operabilidade.

A lógica descritiva parece um opção adequada para representação do conhecimento quando se pensa na combinação de ontologias, lógica e computação para projetos de inter-operabilidade. A extensão proposta, a *DLR* (relações n-árias), parece ser mais condizente com a modelagem do mundo, onde as relações têm diversos componentes. Existem outros dialetos lógicos que podem ser utilizados e verificar qual é o mais adequado é uma decisão de projeto.

A lógica descritiva tem sido usada com sucesso em muitas áreas do conhecimento, inclusive para representação de informações bibliográficas e como suporte para classificação e recuperação em bibliotecas digitais. Um exemplo recente de utilização dessa lógica para construção de ontologias de referencias bibliográficas é apresentada em (WELTY; JENKINS, 2000). A lógica descritiva tem expressividade suficiente para a representação de taxonomias,

comuns em esquemas de classificação de bibliotecas e apresenta vantagens para a classificação e recuperação da informação.

Além de uma solução possível, a abordagem para a inter-operabilidade via utilização de ontologias, pode ser também uma solução mais acessível, no curto e médio prazo, para a recuperação de informação mais eficiente, pela dificuldade em se tratar outros aspectos rapidamente. Esses aspectos dizem respeito à impossibilidade de se obter uma sintaxe uniforme entre diferentes comunidades, e, a dificuldades em tratar questões humanas.

Os modelos representam um mundo complexo, e assim, ferramentas que possibilitem lidar com essa complexidade são essenciais para as atividade de modelagem. O meta-modelo baseado em ontologias foi a proposta apresentada para a integração de fontes que vão compor o ambiente do SABiO. Acredita-se que assim, alcançou-se a contribuição prática proposta no trabalho.

A Ciência da Informação trabalha algumas questões, utilizando ferramentas desenvolvidas por outras áreas. Estudos conjuntos são passíveis de discussão no âmbito da Ciência da Informação, por sua natureza interdisciplinar, e, podem agregar valor a iniciativas centradas no usuário. Ao final desse trabalho, conclui-se que a combinação de estudos de tratamento da informação, através da lógica descritiva e computação, podem compor um tema para trabalhos futuros mais abrangentes. Sem se limitar a abordagem por ontologias, as outras estruturas de organização do conhecimento (esquemas de classificação, *thesaurus*, etc) estudadas nas disciplinas de tratamento da informação, podem ser representadas expressivamente pela lógica descritiva. Estando o conhecimento organizado e expresso formalmente em lógica, ele pode ser manipulado em computadores, de forma a se obterem sistemas digitais cada vez mais “inteligentes” (ou racionais).

O estudo da lógica descritiva e sua forma de representar o conhecimento pode construir uma “ponte” ligando disciplinas de tratamento da informação a disciplinas da inteligência artificial. Essa hipótese foi aqui reforçada pela utilização de ontologias (um tipo de estrutura de organização do conhecimento) para integrar fontes de dados, sendo que a lógica atuou como conexão entre as partes. Acredita-se, dessa forma, que o estudo da relação ontologias e inter-operabilidade possa ser o primeiro passo para uma incursão mais abrangente.

BIBLIOGRAFIA

Referências bibliográficas

- AGUADO, G. et al. (1998). *Ontogeneration: Reusing domain and linguistic ontologies for Spanish text generation?* Disponível em: <<http://delicias.dia.fi.upm.es/WORKSHOP/ECAI98/papers/aguado.ps>>. Acesso em 29 ago. 2002.
- ALBERTAZZI, L.(1996). *Formal and material ontology*. Disponível em: <<http://www.mitteleuropafoundation.it/LA/Papers/Formal%20Material%20Ont.rtf>>. Acesso em: 21 fev. 2002.
- ALLEN, P et al. (2002). *Interoperability between multimedia collections for content and metadata-based searching*. Disponível em: <<http://www.bib.ecs.soton.ac.uk/data/7305/html/index.html>>. Acesso em: 05 out. 2002.
- ALEXAKI, S et al. (2002). *Managing RDF Metadata for Community Webs*. Disponível em: <<http://139.91.183.30:9090/RDF/publications/wcm2000.PDF>>. Acesso em: 11 out. 2002.
- ARENS, Y.; KNOBLOCK, C. A.; SHEN, W. (1996). *Query reformulation for dynamic information integration*. Disponível em: <<http://citeseer.nj.nec.com/arens96query.html>>. Acesso em: 11 maio 2002.
- ARENS, Y.; HSU, C. N.; KNOBLOCK, C. A. (1996). *Query processing in the SIMS information mediator*. Disponível em: <<http://citeseer.nj.nec.com/51138.html>>. Acesso em: 22 fev. 2002.
- ARPA (1995). *Intelligent Integration of Information Reference Architecture*. Disponível em: <http://ise.gmu.edu/I3_Arch/index.html> Acesso em: 15 mar. 2002.
- ARPÍREZ, J. C. et al. (2001). *Web ODE: a Scalable Workbench for Ontological Engineering*. Disponível em: <<http://mkbeem.elibel.tm.fr/paper/KCAP2001-35.pdf>>. Acesso em: 29 jun. 2002.
- BAADER, F.; HOLLUNDER, B. KRIS. (1991). Knowledge representation and inference system. *SIGART Bulletin*, v. 2, n. 3, p. 8-14, 1991.
- BAADER, F. et al. (1992). *Terminological Knowledge representation: a proposal for a terminological logic*. Disponível em: <<http://citeseer.nj.nec.com/baader92terminological.html>>. Acesso em: 22 out. 2002.
- BAADER, F. et al. (2003). *The Description Logic Handbook: Theory, Implementation and Applications*. Appendix Description Logic terminology. Disponível em: <<http://www.ksl.stanford.edu/people/dlm/papers/dlhb-appendix.pdf>>. Acesso em: 15 set. 2002. (a ser publicado em janeiro de 2003).
- BALDONADO, M. COUSINS, S. (1997). *Addressing heterogeneity in the networked information environment*. Disponível em: <<http://citeseer.nj.nec.com/27530.html>>. Acesso em: 27 out. 2001.
- BARRETO, C. M. (1999). *Modelo de metadados para a descrição de documentos eletrônicos na Web*. 190 f. Dissertação (Mestrado em Ciências em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro.
- BATEMAN, J. A. (1996). *Using text structure and text planning to guide text summarization*. Disponível em: <<http://www.ik.fhhannover.de/ik/projekte/Dagstuhl/Abstract/Abstracts/Bateman/Bateman.html>>. Acesso em: 25 maio 2002.

- BATINI, C.; CERI, S.; NAVATHE, S. B. *Conceptual database design, an entity-relationship approach*. Addison Wesley Higher Education, 1992. 470 p.
- BAYARDO, R. J. et al. (1997). *INFOSLEUTH: agent-based semantic integration of information in open and dynamic environments*. Disponível em: <<http://citeseer.nj.nec.com/bayardo97infosleuth.html>>. Acesso em: 30 nov. 2001.
- BAX, P. M. (1997). *Relatório de especificação preliminar da arquitetura*. Disponível em: <<http://www.eb.ufmg.br/bax/Projetos/Sabio/archive.htm>>. Acesso em: 08 abr. 2002.
- BEALE, S., NIRENBURG, S. MAHESH, K. (1995). *Semantic Analysis in the Mikrokosmos Machine Translation Project*. Disponível em: <<http://crl.nmsu.edu/users/sb/papers/thai/thai.html>>. Acesso em: 6 jul. 2002.
- BERGAMASCHI, S.; SARTORI, C. (1997b). *An approach for the extraction of information from heterogeneous sources of textual data*. Disponível em: <<http://sunsite.informatik.rwth-achen.de/Publications/CEUR-WS/Vol-8/paper-1.ps>>. Acesso em 29 out. 2001.
- BERNARAS, A.; LARESGOITI, I.; CORERA, J. (1996). Building and Reusing Ontologies for Electrical Network Applications. *Proceedings of the European Conference on Artificial Intelligence, ECAI/96*, p. 298-302.
- BERTINO, E.; BARBARA, C.; ZARRI, G. P. (1999). *A Conceptual Annotation Approach to Indexing in a Web-Based Information System*. Disponível em: <http://citeseer.nj.nec.com/bertino99conceptual.html>>. Acesso em: 04 dez. 2001.
- BORGIO, S. et al. (1997). *Using a Large Linguistic Ontology for Internet-Based Retrieval of Object-Oriented Components*. Disponível em: <http://citeseer.nj.nec.com/borgio97using.html>>. Acesso em: 20 set. 2002.
- BORST, W.N. (1997). *Construction of Engineering Ontologies*. Phd Thesis. Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 21 abr. 2002.
- BRANCHMAN, R.; SCHMOLZE, J. An overview of the KL-One, Knowledge representation system. *Cognitive Science*, v. 9. n. 2, p. 171-216, 1985.
- BRESSAN, S. (1998). *Semantic Integration of Disparate Information Sources over the Internet using Constraint Propagation*. Disponível em: <<http://context.mit.edu/~steph/cp97/cp97.html>>. Acesso em: 26 jul. 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.
- BROEKSTRA, J. et al. (2001). OIL: a case study in extending RDF-Schema. Disponível em: <<http://www.ontoknowledge.org/oil/oil-rdfs.pdf>>. Acesso em: 25 out. 2002.
- CALVANESE, D. et al. (1998). *Description logic framework for information integration*. Disponível em: <<http://citeseer.nj.nec.com/calvanese98description.html>>. Acesso em: 01 nov. 2002.
- CAMPOS, M. L. A. (1995). *Perspectivas para o estudo da área de representação da informação*. Disponível em: <<http://www.ibict.br/cionline/250296/25029611.pdf>>. Acesso em: 8 jul. 2002.
- CATARCI, T.; LENZERINE, M. (1993). Representing and using interschema knowledge in cooperative information systems. Disponível em: <<http://citeseer.nj.nec.com/catarci93representing.html>>. Acesso em: 19 out. 2002.

- CEN, COMITÉ EUROPÉEN DE NORMALISATION. *Workshop Agreement CWA 13699 – Model for metadata for multimedia information*. Disponível em: <http://www.cenorm.be/iss/cwa_download_area/cwa13699.pdf>. Acesso em: 16 ago. 2002.
- CHANDRASEKARAN, B.; JOHNSON, T. R.; BENJAMINS, V. R. (1999). *Ontologies: what are they? why do we need them?*. Disponível em: <<http://www.computer.org/intelligent/>>. Acesso em 20 set. 2001.
- CHAWATHE, S. et al. (1994). *TSIMMIS Project: integration of heterogeneous Information Sources*. Disponível em: <<http://citeseer.nj.nec.com/chawathe94tsimmis.html>>. Acesso em: 29 abr. 2000.
- CORCHO, O.; FERNÁNDEZ-LÓPEZ, M.; GOMEZ-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: 04 mar. 2002.
- 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-spec/okbc-2-0-3.pdf>>. Acesso em: 09 jun. 2002.
- CORAZZON, R. (2002). *What is ontology?* Disponível em: <http://www.formalontology.it/section_4.htm>. Acesso em: 20 jul. 2002.
- DECKER, S. et al. (2000) *The Semantic Web: the roles of XML and RDF*. Disponível em: <<http://citeseer.nj.nec.com/decker00semantic.html>>. Acesso em: 24 out. 2002.
- DOMINGUE, J. (1998). *Tadzebao and Webonto: Discussing, Browsing and Editing Ontologies on the Web*. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW98/domingue/>>. Acesso em: 30 jun. 2002.
- DOMINGUE, J., MOTTA, E. (1999). *A Knowledge-Based News Server Supporting Ontology-Driven Story Enrichment and Knowledge Retrieval*. Disponível em: <http://kmi.open.ac.uk/~john/papers/planet_onto.pdf>. Acesso em: 22 abr. 2002.
- DOMINGUE, J. et al. (2001) *Supporting Ontology Driven Document Enrichment within Communities of Practice*. Disponível em: <http://eprints.aktors.org/archive/00000012/01/kcap01_john_final.pdf>. Acesso em: 21 jun. 2002.
- 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: 09 jun. 2002.
- DONINI, F. M. et al. (1997). *Reasoning in Description logics*. Disponível em: <<http://citeseer.nj.nec.com/131764.html>>. Acesso em: 24 out. 2002.
- DODDS, D. et al. *Professional XML Metadata*. Wrox Press Inc. Birmingham, 2001
- EUZENAT, J. (1996). *Corporative memory through cooperative creation of knowledge bases and hyper-documents*. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/euzenat/euzenat96b.html>>. Acesso em: 28 set. 2001.
- FARQUHAR, A.; FIKES, R.; RICE, J. (1996). *The Ontolingua Server: A Tool for Collaborative Ontology Construction*. Disponível em: <<http://citeseer.nj.nec.com/farquhar96ontolingua.html>>. Acesso em: 23 set. 2002.
- FENSEL, D. (2001). *Ontologies: a silver bullet for knowledge management and electronic commerce* – Springer – Verlag Berlin Heidelberg, New York, 2001. 137 p.

- FENSEL, D. et al. (2000) *OIL: an ontology infrastructure for the semantic web*. Disponível em: <<http://www.cs.vu.nl/~frankh/postscript/IEEE-IS01.pdf>>. Acesso em: 15 out. 2002.
- FENSEL, D. et al. (1998). *Ontobroker – the very high idea*. Disponível em: <<http://citeseer.nj.nec.com/192612.html>>. Acesso em: 20 dez. 2001.
- FERNÁNDEZ-LÓPEZ et al. (1999). *Building a Chemical Ontology Using Methontology and the Ontology Design Environment*. Disponível em: <<http://www.computer.org/intelligent/ex1999/x1037abs.htm>>. Acesso em: 25 ago. 2001.
- GANDON, F. (2001). *Engineering an Ontology for a Multi-Agents Corporate Memory System*. Disponível em: <<http://citeseer.nj.nec.com/gandon01engineering.html>>. Acesso em: 22 maio 2002.
- GAO, X.; STERLING, L. (2000). Semi-structured data extraction from heterogeneous sources. In: SCHWARTZ, D. G.; DIVITINI, M.; BRASETHVIK, T. *Innovative Internet Information Systems Workshop*. Internet-based organizational memory and knowledge management. New York: Idea Group Publishing. 2000.
- GERMANN, U. (1998). *Making Semantic Interpretation Parser-Independent*. Disponível em: <<http://www.isi.edu/natural-language/people/germann/>>. Acesso em: 11 maio 2002.
- GERSTING, J. L. *Fundamentos matemáticos para a ciência da computação*. 4. ed. Rio de Janeiro: LTC, 2001.
- GOASDOUÉ, F.; LATTES, V.; ROUSSET, M. C. (1998). *The Use of CARIN Language and Algorithms for Information Integration: The PICSEL Project*. Disponível em: <<http://citeseer.nj.nec.com/lattes98use.html>>. Acesso em: 02 ago. 2002.
- GOASDOUÉ, F., REYNAUD, C. (2001). *Modeling information sources for information integration*. Disponível em: <<http://www.lri.fr/~goasdoue/biblio/GoasdoueReynaud-EKAW-1999.pdf.zip>>. Acesso em: 31 mar. 2002.
- GÓMEZ-PÉREZ, A. (2001). *Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases*. Disponível em: <<http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Gomez-Perez1/gomez-perez.pdf>>. Acesso em: 29 jul. 2002.
- GÓMEZ-PÉREZ, A. (1999a). *Some ideas and examples to evaluate ontologies*. Disponível em: <http://www-ksl.stanford.edu/KSL_Abstracts/KSL-94-65.html>. Acesso em: 30 out. 2002.
- GÓMEZ-PÉREZ, A. (1999b). *Knowledge sharing and reuse: ontologies and applications*. Disponível em: <<http://www.ontology.org/main/presentations/madrid/applications.pdf>>. Acesso em: 22 mar. 2002.
- GÓMEZ-PÉREZ, A., BENJAMINS, V. R. (1999) Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods. *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999*.
- GONZALO, J.; PEÑAS, A.; VERDEJO, F. (1999). *Lexical ambiguity and Information Retrieval Revisited*. Disponível em: <<http://citeseer.nj.nec.com/gonzalo99lexical.html>>. Acesso em: 02 fev. 2002.
- GRUBER, T. (1993). *A translation approach to portable ontology specifications*. Disponível em: <http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html>. Acesso em: 04 maio 2001.
- GRUBER, T. (1996). *What is an Ontology?* Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 14 set. 2002.

- GRÜNINGER, M.; FOX, M. S. (1995). *Methodology for the design and evaluation of ontologies*. Disponível em: <<http://citeseer.nj.nec.com/grninger95methodology.html>>. Acesso em: 06 ago. 2001.
- GUARINO, N.; GIARETTA, P. (1995). *Ontologies and KBs, towards a terminological clarification*. Disponível em: <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/KBKS95.pdf>>. Acesso em: 13 jul. 2002.
- GUARINO, N. (1998). *Formal Ontology in information systems*. Disponível em: <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>>. Acesso em: 11 out. 2002.
- GUARINO, N. (1996). *Understanding, building an using ontologies*. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html#Heading4>>. Acesso em: 22 set. 2001.
- GUPTA et al. (2000). *Knowledge based integration of neuroscience data sources*. Disponível em: <<http://www.sdsc.edu/~ludaesch/Paper/ssdbm00.pdf>>. Acesso em: 20 mar. 2002.
- HAAV, H. M.; LUBI, T. L. (2001). *A survey of concept-based information retrieval tools on the web*. Disponível em: <<http://www.science.mii.lt/ADBIS/local2/haav.pdf>>. Acesso em: 31 jul. 2002.
- HASMAN, A. et al. (1999). *ID2.1: Analysis of guideline ontologies*. Disponível em: <<http://new.euromise.org/mgt/repdev/id21.html>>. Acesso em: 29 ago. 2002.
- HEFLIN, J.; HENDLER, J. (2000). *Searching the Web with SHOE*. Disponível em: <<http://www.cs.umd.edu/projects/plus/SHOE/pubs/aiweb2000.pdf>>. Acesso em: 15 jun. 2002.
- HORROCKS, I et al. (2001a). *Reference description of the DAML+OIL ontology markup language*. Disponível em: <<http://www.w3.org/TR/daml+oil-reference>>. Acesso em: 04 nov. 2001.
- HORROCKS, I.; SATTler, U.; TOBIES, S. (1999). *Practical reasoning for expressive description logics*. Disponível em: <<http://citeseer.nj.nec.com/225114.html>>. Acesso em: 21 out. 2002.
- HORROCKS, I. et al. (2001b). *The ontology interchange language OIL*. Disponível em: <<http://public.research.mimesweeper.com/Standards/W3C/RDF-related/oil.pdf>>. Acesso em: 19 out. 2002.
- HWANG, C. H. (1999). *Incompletely and imprecisely speaking: using dynamic ontologies for representing and retrieving information*. Disponível em: <<http://citeseer.nj.nec.com/196692.html>>. Acesso em: 15 dez. 2001.
- JARKE, M.; VASSILIOU, Y. (1997). *Data Warehouse Quality: A Review of the DWQ Project*. Disponível em: <<http://www.dbnet.ece.ntua.gr/~dwq/publications.html>>. Acesso em: 23 jan. 2002.
- JASPER, R.; USCHOLD, M. (1999). *A framework for understanding and classifying ontology applications*. Disponível em: <<http://sem.ucalgary.ca/KSI/KAW/KAW99/papers/Uschold2/final-ont-apn-fmk.pdf>>. Acesso em: 29 set. 2002.
- JIN, L. et al. (1997). *Role Explication of Simulation in Intelligent Training Systems by Training Task Ontology*. Disponível em: <<http://www.ei.sanken.osaka-u.ac.jp/pub/kin/kin-ai97.pdf>>. Acesso em: 28 maio 2002.
- JONES, D.; BENCH-CAPON, T.; VISSER, P. (1998). *Methodologies for ontology development*. Disponível em: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-onto-dev.pdf>>. Acesso em: 17 mar. 2002.

- KALFOGLOU, Y. et al. (2001). *MyPlanet*: an ontology-driven Web-based personalized news service. Disponível em: <<http://eprints.aktors.org/archive/00000010/02/ontoIS01final.pdf>>. Acesso em: 27 fev. 2002.
- KALFOGLOU, Y. (2001). *Deploying Ontologies in Software Design*. PhD Thesis. Disponível em: <<http://www.ecs.soton.ac.uk/~yk1/research.html>>. Acesso em: 21 ago. 2002.
- KARP, R.; CHAUDHRI, V.; THOMERE, J. (1999). *XOL*: an XML-based Ontology Exchange Language. Disponível em: <<http://www.ai.sri.com/~pkarp/xol/xol.html>>. Acesso em: 22 jun. 2002.
- KENT, R. E. (1999). *Conceptual Knowledge Markup Language: The Central Core*. Disponível em: <<http://sern.ucalgary.ca/ksi/kaw/kaw99/papers/Kent1/CKML.pdf>>. Acesso em: 21 jul. 2002.
- KIFER, M.; LAUSEN, G.; WU, J. (1990). *Logical Foundations of Object-Oriented and Frame-Based Languages*. Disponível em: <<http://citeseer.nj.nec.com/kifer90logical.html>>. Acesso em: 01 set. 2002.
- KERHERVÉ, B. (1997). *Models for Metadata or Metamodels for Data?*. Disponível em: <<http://computer.org/conferen/proceed/meta97/papers/bkerherve/bkerherve.html>>. Acesso em: 20 set. 2002.
- KESHAV, R. et al. (1999). *Architecture Integration Elements*. Disponível em: <[http://www.isis.vanderbilt.edu/sdp/Papers/R%20Gamble%20\(Beyond%20Documentation\).doc](http://www.isis.vanderbilt.edu/sdp/Papers/R%20Gamble%20(Beyond%20Documentation).doc)>. Acesso em: 04 mar. 2001.
- KIETZ, J.; MAEDCHE, A.; VOLZ, R. (2000). *A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet*. Disponível em: <http://www.irit.fr/ACTIVITES/EQ_SMI/GRACQ/WSEKAW2000/PAPERS/Maedche.pdf>. Acesso em: 29 set. 2001.
- KNOBLOCK, C. A. et al. (1998). *The Ariadne approach to WEB-Based Information Integration*. Disponível em: <<http://www.isi.edu/info-agents/papers/knoblock98-expert.pdf>>. Acesso em: 27 out. 2001.
- KOHLER, J.; SCHULZE-KREMER, S. (2002). *The semantic metadabase (SEMEDA): ontology based integration of federated molecular biological data sources*. Disponível em: <<http://www.bioinfo.de/isb/gcb01/talks/koehler/main.html>>. Acesso em: 26 ago. 2002.
- LAKATOS, E. M.; MARCONI, M. A. *Fundamentos de metodologia científica*. 3. ed. São Paulo: Atlas, 1991.
- LASSILA, O.; SWICK, R. (1999). *Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation*. Disponível em: <<http://www.w3.org/TR/REC-rdf-syntax/>>. Acesso em: 04 nov. 2001.
- LEGER, A. et al. (2000). *Ontology domain modeling support for multilingual services in e-Commerce: MKBEEM*. Disponível em: <<http://mkbeem.elibel.tm.fr/paper/ecai00-final.pdf>>. Acesso em: 22 dez. 2001.
- LENAT, D. B.; GUHA, R. V. *Building large knowledge-based systems*. Addison-Wesley Publishing Company, Inc. Massachussets, 1990. 372 p.
- LEVY, A. Y.; ROUSSET, M. C. (1996). *CARIN: A Representation Language Integrating Rules and Description Logics*. Disponível em: <<http://www.cs.washington.edu/homes/alon/site/files/ecai96.ps>>. Acesso em 01 set. 2002.
- LÓPEZ, M. F. (1999). *Overview of methodologies for building ontologies*. Disponível em: <<http://www.ontology.org/main/presentations/madrid/analysis.pdf>>. Acesso em: 31 jan. 2002.

LUCK, K. V. et al. (1987). The anatomy of Back system KIT Report 41. *Fachbereich Informatik*, Technische Universität Berlin, Berlin, West-Germany, 1987.

LUKE, S. et al. (1997). *Ontology-based Web agents*. Disponível em: <<http://citeseer.nj.nec.com/luke97ontologybased.html>>. Acesso em: 31 jan. 2002.

MAEDCHE, A. et al. (2001). *Semantic portal: The SEAL approach*. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2001_mit-book.pdf>. Acesso em: 15 abr. 2002.

MAEDCHE, A. (2000). *A machine-learning perspective for the semantic web – Position paper*. Disponível em: <<http://www.semanticweb.org/SWWS/program/position/soi-maedche.pdf>>. Acesso em: 5 set. 2002.

MAEDCHE, A. et al. (2000). *Representation Language-Neutral Modeling of Ontologies*. Disponível em: <http://www.ontoprise.de/download/ontoedit_paper.pdf>. Acesso em: 21 out. 2002.

MADCHE, A.; STAAB, S.; STUDER, R. (2000). *Ontology-Oriented Information Extraction and Integration*. Disponível em: <<http://www.aifb.uni-karlsruhe.de/~sst/Research/Publications/ssarbruecken99.pdf>>. Acesso em: 06 mar. 2002.

MAHALINGAM, K.; HUHNS, M. N. (1997). *An Ontology Tool for Query Formulation in an Agent-Based Context*. Disponível em: <<http://www.cse.sc.edu/research/cit/CoopIS/papers97/coopis97-kuha.html>>. Acesso em: 20 out. 2002.

MARTIN, P. H.; EKLUND, P. (2001). Large-scale cooperatively-built heterogeneous KBs. In: *ICCS'01, 9th INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES*. Disponível em: <<http://meganesia.int.gu.edu.au/~phmartin/WebKB/doc/papers/iccs01/>>. Acesso em: 6 set. 2002.

MENA, E. et al. (1996). *OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*. Disponível em: <<http://citeseer.nj.nec.com/mena96observer.html>>. Acesso em: 11 jul. 2002.

MENDES, R. D. (1997). *Inteligência artificial: sistemas especialistas no gerenciamento da informação*. Disponível em: <<http://www.ibict.br/cionline/260197/26019705.pdf>>. Acesso em: 14 out. 2002.

MIZOGUCHI, R.; VANWELKENHUYSEN, J., IKEDA, M. (1995). *Task ontology for reuse of problem solving knowledge*. Disponível em: <http://citeseer.nj.nec.com/context/24780/0>. Acesso em: 04 mar. 2001.

MOF, MetaObject Facility. *Models and Metamodels*. Disponível em: <<http://cs.ua.edu/630/ReadingsAndMaterials/MOF%2000-04-03.pdf>>. Acesso em: 19 out. 2002.

MORENO, A. O.; HERNÁNDEZ, C. P. (2000). *Reusing the Mikrokosmos Ontology for Concept-Based Multilingual Terminology Databases*. Disponível em: <<http://webdeptos.uma.es/filifa/personal/amoreno/research/pubs.htm>>. Acesso em: 30 mar. 2002.

MOTTA, E.; BUCKINGHAM-SHUM, S.; DOMINGUE, J. (2000). *Ontology-Driven Document Enrichment: Principles, Tools and Applications*. Disponível em: <<http://kmi.open.ac.uk/projects/scholonto/docs/OnDocEnrich-IJHCS-2000.pdf>>. Acesso em: 10 jan. 2002.

NEBEL, B. (1998). *Principles of KR and reasoning*. Apresentação em PowerPoint. Disponível em: <<http://www.informatik.uni-freiburg.de/~nebel/index.html?>>. Acesso em: 8 ago. 2002.

- NEBEL, B. (2001). *Logics for knowledge representation*. Disponível em: <<ftp://ftp.informatik.uni-freiburg.de/documents/papers/ki/nebel-iesbs-01.pdf>>. Acesso em: 02 out. 2002.
- NECHES, R. et al. (1991). *Enabling technology for knowledge sharing*. Disponível em: <<http://www.isi.edu/isd/KRSharing/vision/AIMag-small.html>>. Acesso em: 20 mar. 2002.
- NEODOBITY, W. (1986). Classification, artificial intelligence and cognitive psychology. In: GAUL, W.; SCHADER, M. *Classification as a tool for research*. Amsterdam: Elsevier, 1986. p. 245-352.
- NOY, F. N.; GUINNESS, D. L. (2001). *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: 04 maio 2001.
- NOY, N. F.; MUSEN, M. A. (2000). *PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment*. Disponível em: <http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-2000-0831.pdf>. Acesso em: 22 jan. 2002.
- NOY, N. F.; FERGERSON, R. W.; MUSEN, M. A. (2000). *The knowledge model of Protege-2000: Combining interoperability and flexibility*. Disponível em: <http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-2000-0830.pdf>. Acesso em: 31 ago. 2002.
- OGATA, N. (2001). *A formal ontology discovery from web documents*. Disponível em: <<http://lc301.lang.osaka-u.ac.jp/maebashi5.pdf>>. Acesso em: 11 maio 2002.
- OGURA, K.; BOND, F.; OYAMA, Y. (1999). *ALT-J/M: A prototype Japanese-to-Malay Translation System*. Disponível em: <<http://www.kecl.ntt.co.jp/icl/mtg/members/bond/pubs/1999-summit-jm.pdf>>. Acesso em: 16 mar. 2002.
- OMELAYENKO, B. (2001). *Integration of product ontologies for B2B marketplaces: a preview*. Disponível em: <[http://www.acm.org/sigs/sigecom/exchanges/volume_2_\(01\)/2.1-Omelayenko.pdf](http://www.acm.org/sigs/sigecom/exchanges/volume_2_(01)/2.1-Omelayenko.pdf)>. Acesso em: 22 jun. 2002.
- OUKSEL, A. M.; SHETH, A. (1999). *Semantic Interoperability in Global Information Systems*. Disponível em: <<http://sirius.cs.ucdavis.edu/teaching/289F-SQ00/reading.html>>. Acesso em: 22 jan. 2002.
- OZKURAL, E. (2001). *Ontology Tools for Repositories on Internet*. Disponível em: <<http://borg.cs.bilkent.edu.tr/~exa/ontology/>>. Acesso em: 07 set. 2002.
- PACHECO, R. C. S.; KERN, V. M. (2001). *Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia, e Kern*. Disponível em: <<http://www.ibict.br/cionline/300301/3030801.pdf>>. Acesso em: 22 abr. 2002.
- PAEPCKE, A. (1996). Digital libraries: searching is not enough. *D-Lib Magazine*, may 1996. Disponível em: <<http://www.dlib.org/dlib/may96/stanford/05paepcke.html>>. Acesso em: 03 fev. 2002.
- PALEY, J. L. S.; KARP, P. (1997). *A Generic Knowledge base Browser and Editor*. Disponível em: <ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-97-11.pdf>. Acesso em: 09 out. 2002.
- PATEL-SCHNEIDER, P. F.; SWAROUT, B. (1993). *Description-logic knowledge representation system specification from the KRSS Group of the ARPA Knowledge Effort*. Disponível em: <http://www-db.research.bell-labs.com/FwrmrcrhFkHw_ers/krss-spec.ps>. Acesso

em: 29 out. 2002.

PINTO, P. R. M. *Introdução à lógica simbólica*. Belo Horizonte: Ed. UFMG, 2001.

PREECE, A.; HUI, K. (1999). The KRAFT architecture for knowledge fusion and transformation. Disponível em: <<http://citeseer.nj.nec.com/364207.html>>. Acesso em: 21 jan. 2002.

RECTOR et al. (1997). The Grail Concept Modeling Language for medical terminology. Disponível em: <<http://potato.cs.man.ac.uk/papers/aim97.pdf>>. Acesso em: 04 nov. 2001.

ROBREDO, J. Documentação de hoje e amanhã. 2. ed .rev. e amp. Brasília: Ed. Autor, 1986. p. 201-263.

RÖSNER, D.; STEDE, M. (1994). *Techdoc*: Multilingual generation of online and offline instructional text online and offline instructional text. Disponível em: <<http://acl.ldc.upenn.edu/A/A94/A94-1044.pdf> >. Acesso em: 22 jul. 2002.

SARACEVIC, T. Ciência da Informação: Origem, evolução e relações. *Perspec. Ci. Inf.*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan. / jun. 1996.

SAYÃO, L. F. (2001). *Modelo teóricos em Ciência da Informação* – abstração e método científico. Disponível em: <<http://dois.mimas.ac.uk/DoIS/data/Articles/juljqbfchy:2001:v:30:i:1:p:1847.html>>. Acesso em: 22 set. 2002.

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.

SELIGMAN, L. (1995). *Metrics for accessing heterogeneous data*: Is there any hope? Disponível em: <http://www.mitre.org/pubs/data_mgt/Papers/vldb95panel.pdf>. Acesso em: 10 mar. 2001.

SKUCE, D. (1995). *CODE4*: a Unified System for Managing Conceptual Knowledge. Disponível em: <<http://citeseer.nj.nec.com/skuce95code.html>>. Acesso em: 02 set. 2002.

SKUCE, D. (1996). *IKARUS*: Intelligent knowledge acquisition and retrieval universal system. Disponível em: <<http://www.csi.uottawa.ca/~kavanagh/Ikarus/Ikarus4.html>>. Acesso em: 04 out. 2002.

STAAB, S. et al. (2001). *Knowledge Processes and Ontologies*. Disponível em: <<http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications.html>>. Acesso em: 22 jul. 2001.

STOCK, O. (1996) *GIST* – Final Report to European Community Linguistic Research and Engineering Programmed. Disponível em: <<http://ecate.itc.it:1025/projects/gist/deliverables/Final-report-new.ps.gz>>. Acesso em: 16 out. 2002.

STUCKENSCHMIDT, H., WACHE, H. (2000). *Context modeling and transformation for semantic interoperability*. Disponível em: <<http://citeseer.nj.nec.com/stuckenschmidt00context.html>>. Acesso em: 02 dez. 2001.

STUCKENSCHMIDT, H. et al. (2000). *Catalogue integration*: a case study in ontology-based semantic translation. Disponível em: <<http://citeseer.nj.nec.com/stuckenschmidt00catalogue.html>>. Acesso em: 11 ago. 2002.

STUMME, G.; MAEDCHE, A.. (2001). *FCA-Merge*: Bottom-Up Merging of Ontologies. Disponível em: <<http://citeseer.nj.nec.com/stumme01fcamerge.html>>. Acesso em: 12 dez. 2001.

- SOWA, J. F. (1999a). *Preface to Knowledge representation*. Disponível em: <<http://users.bestweb.net/~sowa/krbook/krpref.htm>>. Acesso em: 16 jul. 2002.
- SOWA, J. F. (1999b). *Building, sharing and merging ontologies*. Disponível em: <<http://www.jfsowa.com/ontology/ontoshar.htm>>. Acesso em: 30 maio 2002.
- SUBRAHMANIAN, V. et al. (1992). *Hermes: heterogeneous reasoning and mediator system*. Disponível em: <<http://www.cs.umd.edu/projects/hermes/overview/paper/index.html>>. Acesso em: 23 fev. 2002.
- SWARTOUT, B. et al. (1996). *Toward Distributed Use of Large-Scale Ontologies*. Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em: 22 jul. 2001.
- TEICH, E.; BATEMAN, A. J. (1993). *Towards the application of text generation in an integrated publication system*. Disponível em: <<http://www.darmstadt.gmd.de/publish/komet/pubs/nlgw94.ps.gz>>. Acesso em: 26 out. 2002.
- TENNISON, J.; SHADBOLT, N. R. (1998). *APECKS: a Tool to Support Living Ontologies*. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW98/tennison>>. Acesso em: 30 maio 2002.
- TOMASIC, A.; RASCHID, L.; VALDURIEZ, P. (1995). *Scaling heterogeneous distributed databases and the design of Disco. Proceedings of the 16th International Conference on Distributed Computing Systems*, Hong Kong, 1995.
- UK Interoperability Focus. Disponível em: <<http://www.ukoln.ac.uk/interop-focus>>. Acesso em: 25 mar. 2002.
- USCHOLD, M.; KING, M. (1996). *Building Ontologies: Towards a Unified Methodology*. Disponível em: <<http://citeseer.nj.nec.com/uschold96building.html>>. Acesso em: 04 ago. 2001.
- USCHOLD, M.; GRUNINGER, M. (1996b). *Ontologies: principles, methods an applications*. Disponível em: <<http://citeseer.nj.nec.com/uschold96ontologie.html>>. Acesso em: 21 set. 2002.
- VAN HEIJIST, G.; SCHREIBER, A. T.; WIELINGA, B. J. (2002). *Using Explicit Ontologies in KBS Development*. Disponível em: <<http://ksi.cpsc.ucalgary.ca/IJHCS/VH/VH1.html>>. Acesso em: 04 maio 2001.
- 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.
- WACHE, H. et al. (2001). *Ontology-based integration of information – a survey of existing approaches*. Disponível em: <<http://citeseer.nj.nec.com/wache01ontologybased.html>>. Acesso em: 01 maio 2002.
- WATSON, A. (2000). *Workshop Object Management Group – OMG: Creating Interoperability*. Disponível em: <<http://www.internet2.edu/health/20000710-SURA-AWatson.pdf>>. Acesso em: 25 mar. 2002.
- WEIHAI, Y. (2002). *Middleware Seminar*. Disponível em: <<http://www.cs.uit.no/studier/kurs/d312/info/2001h/forelesning/middleware.pdf>>. Acesso em: 25 mar. 2002.
- WELTY, C.; GUARINO, N. (2000). *Supporting Ontological Analysis of Taxonomic Relationships*. Disponível em: <<http://citeseer.nj.nec.com/452329.html>>. Acesso em: 02 jul. 2002.
- WIEDERHOLD, G. (1999). *Mediation to deal with heterogeneous data sources*. Disponível em: <<http://www-db.stanford.edu/pub/gio/1999/Interopdocfigs.html>>. Acesso em: 21 fev. 2002.

WIEDERHOLD, G. (1992). *Mediation in the Architecture of Future Information Systems*. Disponível em: <http://www-db.stanford.edu/LIC/mediator.html>>. Acesso em: 15 jan. 2002.

ZDRAHAL, Z. et al. (2000). Sharing engineering design knowledge in a distributed environment. Disponível em: <http://kmi.open.ac.uk/projects/enrich/BIT-sharing-design.pdf>>. Acesso em: 8 set. 2002.

Bibliografia complementar

ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. *Data On the Web – from relations to semistructured data and XML*. San Francisco: Morgan Kaufman, 2000. 257 p.

ARENS et al. (1993). *Retrieving and Integrating data from multiple information sources*. Disponível em: <http://citeseer.nj.nec.com/arens93retrieving.html>>. Acesso em: 27 out. 2001.

ARMS, Y. W. *Digital libraries*. Cambridge: The MIT, 2000. 287 p.

BALDONADO, M.; WINOGRAD, T. (1996). *Sense Maker: an information-exploration interface supporting the contextual evolution of user's interests*. Disponível em: <http://www.psrg.lcs.mit.edu/projects/inforadar/p11-baldonado.pdf>>. Acesso em: 28 out. 2001.

BALDONADO, M.; CHANG, C. C.; GRAVANO, L. (1997). *Metadata for Digital Libraries: architecture and design rationale*. Disponível em: <http://www-diglib.stanford.edu/diglib/pub/slides/dl97-metadata-slides/>>. Acesso em: 27 out. 2001.

BERGAMASCHI, S.; CASTANO, S.; VINCINI, M. (1999) *Semantic Integration of semi structured and structured data sources*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi99semantic.html>>. Acesso em: 29 out. 2001.

BERGAMASCHI, S. (1997). *Extraction of information from highly heterogeneous source of textual data*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi97extraction.html>>. Acesso em: 29 out. 2001.

BERGAMASHI et al. (1997). *Object wrapper: an Object-Oriented interface for relational databases*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi97object.html>>. Acesso em: 29 out. 2001.

BERGAMASHI, S. et al. (1997). *Intelligent techniques for the extraction and integration of heterogeneous information*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi99intelligent.html>>. Acesso em: 9 out. 2001.

BERGAMASCHI et al. (1998). *An intelligent approach to information integration*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi98intelligent.html>>. Acesso em: 29 out. 2001.

BERGAMASCHI, S. et al. (1998a). *Exploiting schema knowledge for the integration of heterogeneous sources*. Disponível em: <http://citeseer.nj.nec.com/bergamaschi98exploiting.html>>. Acesso em: 29 out. 2001.

BISKUP, J.; EMBLEY, D. M. (2001). *Extracting information from heterogeneous information sources using ontologically specified target views*. Disponível em: <http://www.deg.byu.edu/papers/int.pdf>>. Acesso em: 20 nov. 2001.

BORGHOFF et al. (1997). *Agent-Based Document Retrieval for the European Physicists: A Project Overview*. Disponível em: <http://www.xrce.xerox.com/research/ct/projects/cbkb/home.html>>. Acesso em: 20 nov. 2001.

BOWMAN, C. M. et al. (1995). *The Harvest Information Discovery and Access System*. Disponível em: <http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/schwartz.harvest/>

schwartz.harvest.html#papers>. Acesso em: 20 nov. 2001.

BRESSAN, C. K. et al. (1998). *Demonstration of the Context Interchange Mediator Prototype*. Disponível em: <<http://context.mit.edu/~coin/publications/>>. Acesso em: 20 nov. 2001.

BRITTON, C. *IT Architectures an Middleware*. Boston: Addison-Wesley, 2001. 296 p.

CHANDRASEKARAN, B.; JOSEPPHSON, J. R.; BENJAMINS, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, v. 14, n. 1, p. 20-26, 1999.

CHAITANYA, B. et al. (1999). *XML-Based Information Mediation with MIX*. Disponível em: <<http://www.db.ucsd.edu/publications/vamp.pdf>>. Acesso em: 27 out. 2001.

CHANG, C. C. K.; GARCIA-MOLINA, H.; PAEPCKE, A. (1996). *Predicate rewriting for translating boolean queries in a heterogeneous information system*. Disponível em: <<http://citeseer.nj.nec.com/chang96predicate.html>>. Acesso em: 30 ago. 2001.

CHUANG, J. (1999). *Communications middleware*. Disponível em: <<http://www.sims.berkeley.edu/courses/is206/f99/slides/is206.99.10.21.1up.pdf>>. Acesso em: 03 dez. 2001.

CODY, W. F. et al. (1995). *Querying Multimedia Data from Multiple Repositories by Content: the Garlic Project*. Disponível em: <<http://www.almaden.ibm.com/cs/garlic/homepage.html>>. Acesso em: 03 dez. 2001.

COHEN, W. (1998). *Integration of heterogeneous databases without common domains using queries based on textual similarity*. Disponível em: <<http://citeseer.nj.nec.com/cohen98integration.html>>. Acesso em: 03 dez. 2001.

COUSINS, S. et al. (1997). *Towards wide-area distributed interfaces*. Disponível em: <<http://dbpubs.stanford.edu/pub/1997-67>>. Acesso em: 03 dez. 2001.

CHAMBERLAIN, D.; ROBIE, J.; FLORESCU, D. (2000). *QUILT: An XML query language for heterogeneous data sources - (199X)*. Disponível em: <<http://citeseer.nj.nec.com/chamberlin00quilt.html>>. Acesso em: 03 dez. 2001.

ERDMANN, M.; STUDER, R. (1999). *Ontologies as conceptual models for XML documents*. Disponível em: <<http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Erdmann1/erdmann.pdf>>. Acesso em: 05 dez. 2001.

FANKHAUSER, P. et al. (1998). *Experiences in Federated Databases: From IRO-DB to MIRO-Web*. Disponível em: <<http://www.acm.org/sigmod/vldb/conf/1998/p655.pdf>>. Acesso em: 05 dez. 2001.

FAULSTICH, L. C. (1998). *Using Graph Transformation Techniques for Integrating Information from the WWW*. Disponível em: <<http://www.inf.fu-berlin.de/~faulstic/research/HyperView/>>. Acesso em: 05 dez. 2001.

FENSEL, D. et al. (1999). *A Unifying Framework for the Knowledge Web*. Disponível em: <<http://www.cs.vu.nl/~dieter/ftp/spool/upml.oil.pdf>>. Acesso em: 04 nov. 2001.

FLORESCU, D.; LEVY, A.; MENDELZON, A. (1998). *Database techniques for the World Wide Web: a survey*. Disponível em: <<http://citeseer.nj.nec.com/florescu98database.html>>. Acesso em: 15 mar. 2001.

FRANCONI, E. (1997). *Description Logis and Logics*. Disponível em: <<http://citeseer.nj.nec.com/74961.html>>. Acesso em: 28 jun. 2002.

FRANK, A. (1997). *Ontology, a consumer's point of view*. Disponível em: <http://www.dpi.inpe.br/geopro/referencias/ontology_af.pdf>. Acesso em: 15 mar. 2001.

FREDERKING, R. et al. (1997). *Integrating Translations from Multiple Sources within the Pangloss Mark III Machine Translation*. Disponível em:

<<http://www-2.cs.cmu.edu/~ref/papers.html>>. Acesso em: 30 set. 2002.

GARDARIN, G.; SHA, F.; NGOC, T. D. (1999). *XML-based components for federating multiple heterogeneous data sources*. Disponível em:

<<http://www.e-xmlmedia.fr/White-paper%20HDI.pdf>>. Acesso em: 06 dez. 2001.

GENESERETH, M. R.; KELLER, A. M.; DUSCHKA, O. M. (1997). *Infomaster: An Information Integration System*. Disponível em:

<<http://infomaster.stanford.edu/infomaster-info.html>>. Acesso em: 06 dez. 2001.

GEISLER, E. (1999). *The metrics of technology evaluation: where we stand and where we should go from here*. Disponível em: <<http://www.stuart.iit.edu/workingpapers/technology>>. Acesso em: 6 dez. 2002.

GÓMEZ-PÉREZ, A. (1999). *Ontological engineering*. Disponível em:

<<http://www.ontology.org/main/presentations/madrid/theoretical.pdf>>. Acesso em: 04 ago 2001.

GRAHAM, T. Conceptions of library and information Science. *Proceedings of the international Conference held for the celebration of 20th Anniversary of the Department of Information Studies*, University of Tampere, Finland, 26-28 August 1991, p. 309.

GRAVANO, L.; CHANG, C. C.; GARCIA-MOLINA, H. (1996). *STARTS: Stanford Proposal for Internet Meta-Searching*. Disponível em:

<<http://www-db.stanford.edu/pub/gravano/1996/sigmod97.ps>>. Acesso em: 05 dez. 2001.

GRAVANO, L.; GARCIA-MOLINA, H.; TOMASIC, A. GIOSS: text-source discovery on the Internet. *ACM TODS*, v. 24, n. 2, p. 229-264, 1999.

GUARINO, N. (1995). *Formal ontology, conceptual analysis and knowledge representation*. Disponível em: <<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FormOntKR.pdf>>. Acesso em: 30 maio 2002.

HAWKINS, T. D. (2001). *Bibliometrics of electronic journals in information science*. Disponível em: <<http://informationr.net/ir/7-1/paper120.html>>. Acesso em: 6 dez. 2002.

HUFFMAN, S. (1995). Heuristic joins to integrate structured heterogeneous data. Disponível em: <<http://citeseer.nj.nec.com/huffman95heuristic.html>>. Acesso em: 12 dez. 2001.

INTEROPERABILITY reference guide – Industry Advisory Council / Interoperability committee. Disponível em:

<http://www.cio.gov/Documents/interoperability_ref_guide_Oct_1997.html>. Acesso em: 20 maio 2002.

IVES, G. Z. et al. (1999). *An Adaptive Query Execution System for Data Integration*. Disponível em: <http://data.cs.washington.edu/papers/iffiw_sigmod99.ps>. Acesso em: 05 jan. 2002.

JACOBS, D. (2001). *A bibliometric study of the publication patterns of scientists in South Africa 1992-96, with particular reference to status and funding*. Disponível em: <<http://informationr.net/ir/6-3/paper104.html>>. Acesso em: 6 dez. 2002.

JEONG, E.; CHUN-NAN, H. (2001). *Induction of integrated view for xml data with heterogeneous DTDs*. Disponível em:

<<http://www.iis.sinica.edu.tw/~chunnan/DOWNLOADS/cikm2001.pdf>>. Acesso em: 10 ago. 2001.

KAMIYA, K.; ROSCHEISEN, M.; WINOGRAD, T. (2001). *Grassroots: a system providing a uniform framework for communicating, structuring, sharing information and organizing people*. Disponível em: <<http://citeseer.nj.nec.com/kamiya96grassroots.html>>. Acesso em: 27 out. 2001.

KUSHMERICK, N.; WELD, D.; DOORENBOS, R. (1997). *Wrapper induction for information*

extraction. Disponível em: <<http://citeseer.nj.nec.com/kushmerick97wrapper.html>>. Acesso em 27 out. 2001.

LAENDER, A. et al. (2002). *A brief survey of web data extraction tools*. Disponível em: <<http://www.lbd.dcc.ufmg.br/~alti/pubs/sigrec02.ps.gz>>. Acesso em: 10 jul. 2001.

LALANDE, A. *Vocabulário técnico e crítico da filosofia*. 2. ed. São Paulo: Martins Fontes, 1996.

LEVY, A. (1999). *The information Manifold approach to data integration*. Disponível em: <<http://citeseer.nj.nec.com/hammer99information.html>>. Acesso em: 12 jan. 2002.

LINTHICUM, D. S. *Enterprise Application Integration*. Boston: Addison-Wesley, 2000. 375 p.

LIU, L.; PU, C. (1995). *The DIOM Approach to Large-scale Interoperable Database Systems*. Disponível em: <<http://citeseer.nj.nec.com/96432.html>>. Acesso em: 28 out. 2001.

LIU, L.; CALTON, P.; HAN, W. (2000). *XWRAP: an XML-enabled wrapper construction system for web information sources*. Disponível em: <<http://citeseer.nj.nec.com/liu00xwrap.html>>. Acesso em: 28 out. 2001.

LOOM GUIDE 91. *Loom User's Guide for Loom version 1.4. ISX Corporation*. Disponível em: <<http://www.isi.edu/isd/LOOM/documentation/LOOM-DOCS.html>>. Acesso em: 6 set. 2002.

MARC Standards – *Understanding MARC bibliographic*. Disponível em: <<http://lcweb.loc.gov/marc/umb/>>. Acesso em: 27 jul. 2001.

MARAZAKIS, M.; PAPADAKIS, D.; NICOLAU, C. (1997). *Aurora: an architecture for dynamic and adaptative work sessions in open environments*. Disponível em: <<http://citeseer.nj.nec.com/268386.html>>. Acesso em: 08 nov. 2001.

MANOLESCU, I.; FLORESCU, D.; KOSSMANN, D. (2001). *Answering XML queries over heterogeneous data sources*. Disponível em: <http://www.cs.ruu.nl/docs/vakken/gd/answering_xml_queries.pdf>. Acesso em: 05 set. 2001.

MECCA, G. et al. (1998). *The Araneus Web-Base Management System*. Disponível em: <<http://www.dia.uniroma3.it/Araneus/publications/sigmod98.ps>>. Acesso em: 08 nov. 2001.

METHODOLOGY of the definition abstracts from the ISO standards - Abstract from the ISO standard 704: Principles and methods of terminology. Disponível em: <http://www.ohi.shom.fr/Dhydro/Html/site_edition/Help_HTML/isodef.html#i>. Acesso em: 20 out. 2002.

MERRIAM-WEBSTER; GOVE, P. B. *Webster's Third New International Dictionary*. Unabridged. New York: Merriam-Webster, 2002. 2783 p.

MILLER, R. (1998). *Using Schematically heterogeneous structures*. Disponível em: <<http://citeseer.nj.nec.com/miller98using.html>>. Acesso em: 05 set. 2001.

NODINE, M.; BOHRER, W.; NGU, A. H. (1998). *Semantic Brokering over Dynamic Heterogeneous Data Sources in InfoSleuth*. Disponível em: <<http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/woelk/woelk.html>>. Acesso em: 20 mar. 2002.

PAEPCKE, A.; CHNAG, C. C. K.; GARCIA-MOLINA, H.; WINOGRAD, T. (1998). *Interoperability for digital libraries: problems and directions*. Disponível em: <<http://citeseer.nj.nec.com/119879.html>>. Acesso em: 03 fev. 2002.

PAEPCKE, A. et al. (1996). *Towards interoperability in digital libraries overview and selected highlights of the Stanford Digital Library Project*. Disponível em: <<http://www-diglib.stanford.edu/diglib/pub/reports/>>. Acesso em: 03 fev. 2002.

PAEPCKE et al. (2000). *Search Middleware and Simple Digital Library Interoperability Protocol (long version)*. Disponível em: <<http://dbpubs.stanford.edu/pub/2000-52>>. Acesso em: 03 fev. 2002.

- PANTI, M.; SPALAZZI, L.; GIRETTI, A. (2000). *A case-based approach to information integration*. Disponível em: <<http://citeseer.nj.nec.com/461630.html>>. Acesso em: 27 out. 2001.
- PAPAKONSTANTINOY, Y.; GARCIA-MOLINA, H.; WIDOW, J. (1995). *Object exchange across heterogeneous information sources*. Disponível em: <<http://citeseer.nj.nec.com/papakonstantinou95object.html>>. Acesso em: 27 maio 2001.
- PEIG, E.; DELGADO, J.; PÉREZ, I. (2001). *Metadata Interoperability e meta-search on the web*. Disponível em: <<http://www.nii.ac.jp/dc2001/proceedings/product/paper-37.pdf>>. Acesso em: 13 maio 2002.
- QUINOT, T.; KORDON, F.; PAUTET, L. (2002). *From functional to architectural analysis of a middleware supporting interoperability across heterogeneous distribution models*. Disponível em: <<http://www.infres.enst.fr/~quinot/publis/doa01.ps>>. Acesso em: 25 mar. 2002.
- ROSCHISEN, M. et al. (1997). *The Stanford InfoBus and Its Service Layers*. Disponível em: <<http://citeseer.nj.nec.com/100891.html>>. Acesso em: 27 out. 2001.
- ROSENBLUM, D. (2001). *Interoperability & middleware*. Disponível em: <<http://www1.ics.uci.edu/~dsr/old-courses/ics221w2001/221-20010208.pdf>>. Acesso em: 27 out. 2001.
- ROWLEY, J. *Organizing Knowledge: An introduction to information retrieval*. 2. ed. Vermont: Gower, 1996. 510 p.
- RUH, W.; MAGINNIS, F.; BROWN, W. J. *Enterprise Application Integration – a Wiley tech brief*. Wiley Computer Publishing, 2001. 211 p.
- RUSSEL, S.; NORVIG, P. *Artificial Intelligence – a modern approach*. New Jersey: Prentice Hall, 1995. 912 p.
- SAHUGUET, A.; AZAVANT, F. (1999). *WysiWyg Web Wrapper factory (W4F)*. Disponível em: <<http://db.cis.upenn.edu/DL/WWW8/>>. Acesso em: 28 out. 2001.
- SATTLER, K.; SAAKE, G. (1999). *Supporting Information Fusion with Federated Database Technologies*. Disponível em: <<http://fusion.cs.uni-magdeburg.de/infuse/>>. Acesso em: 28 out. 2001.
- SEYMORE, K.; MCCALLUN, A.; ROSENFELD, R. (1999). *Learning Hidden Markov model structured for information extraction*. Disponível em: <<http://citeseer.nj.nec.com/seymore99learning.html>>. Acesso em: 02 out. 2001.
- SHAW, M.; GARLAN, D. (2002). *Formulations and formalisms in software architecture*. Disponível em: <<http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/vit/ftp/pdf/ProgCodif.pdf>>. Acesso em: 11 mar. 2002.
- SHUM, S. B.; MOTTA, E.; DOMINGUE, J. (2000). *ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse*. Disponível em: <<http://citeseer.nj.nec.com/buckinghamshum00scholonto.html>> Acesso em: 28 mar. 2002.
- STUDER, B.; FENSEL, D. (1998). *Knowledge Engineering: principles and methods*. Disponível em: <<http://citeseer.nj.nec.com/225099.html>>. Acesso em: 15 abr. 2002.
- TENNANT, R. (1998). *Interoperability: the holy grail*. Disponível em: <www.ljdigital.com>. Acesso em: 03 nov. 2001.
- UNIVERSITY of Leeds. *Model Markov Tutorial*. Disponível em: <http://www.comp.leeds.ac.uk/scs-only/teaching-materials/HiddenMarkovModels/html_dev/main.html>. Acesso em: 10 ago. 2001.
- WELTY, C.; JENKINS, J. (2000). *Untangle: a new ontology for card catalog systems*. Disponível em: <<http://untangle.cs.vassar.edu/>>. Acesso em: 16 dez. 2002.

WIEDERHOLD, G. (1994). *Interoperation, mediation and ontologies*. Disponível em: <<http://www-db.stanford.edu/pub/gio/1994/medont.ps>>. Acesso em 20 mar. 2002.

ANEXOS

Anexo A – Algumas definições e símbolos da lógica descritiva

A1) Principais operadores da lógica descritiva

Construtor	Sintaxe	Exemplo	Semântica obtida pela interpretação
Conceito atômico	A	<i>Humano</i>	$A^I \zeta \Delta^I$
Papel atômico	R	<i>Tem-criança</i>	$R^I \zeta \Delta^I \times \Delta^I$

Figura A1 – Exemplos de construtores para a lógica descritiva

Construtor	Sintaxe	Exemplo	Semântica obtida pela interpretação
Conjunção	$C \sqcap D$	Humano $\sqcap$ Masculino	$C^I \cap D^I$
Disjunção	$C \sqcup D$	Doutor $\sqcup$ advogado	$C^I \cup D^I$
Negação	$! C$	! masculino	Δ^I / c^I
Restrição existencial	$E R.C$	Etem-criança.masculino	$\{x \mid \exists y. \langle x, y \rangle \in R^I \wedge y \in C^I\}$
Restrição valor	$AR.C$	Atem-criança.doutor	$\{x \mid \forall y. \langle x, y \rangle \in R^I \rightarrow y \in C^I\}$
Restrições	$\bigcup_n R.C$	$\bigcup_1$ pais.feminino	$\{x \mid \bigvee_n \{y. \langle x, y \rangle \in R^I \wedge y \in C^I\}\}$
Numero qualificado	$\bigwedge_n R.C$	$\bigwedge_3$ crianças.feminino	$\{x \mid \bigwedge_n \{y. \langle x, y \rangle \in R^I \wedge y \in C^I\}\}$

Figura A2 – Exemplos de construtores para a lógica descritiva

A2) Sobre a noção de interpretação

Outro conceito importante que deve ser introduzido é o de *interpretação*, simbolizado por I , que consiste em um *domínio* e uma *função* que atribuem um valor da função a cada *nome de conceito*, *nome de papel* e *nome de indivíduo*. Exemplos da semântica dada por uma interpretação $I=(\Delta^I, .^I)$ são apresentados nas Figuras A1 e A2 (NEBEL, 1998). Na notação utilizada a construção $x.^I$ é igual a $I(x)$, ou seja, uma função de x (por exemplo, $1.^I = I(1) = 1$).

A idéia de semântica no universo da lógica descritiva consiste de objetos, funções sobre objetos e relações sobre objetos. Símbolos de funções são mapeados para funções, símbolos de predicados são mapeados para relações. A semântica é representada por uma *interpretação*, conforme definida anteriormente, $I=(\Delta^I, .^I)$ onde Δ^I é um conjunto arbitrário não vazio (domínio da interpretação) e $.^I$ uma função (função de interpretação) que:

- Mapea símbolos de funções “n-árias” para funções “n-árias” no domínio da interpretação f^I
 $\in [(\Delta^I)^n \rightarrow \Delta^I]$;
- Mapea símbolos de predicados n-ários para relações “n-árias” : $P^I \subseteq (\Delta^I)^n$;
- Mapea constantes (conceitos de aridade zero) para objetos.

A figura a seguir torna mais claro o conceito de interpretação :



Figura A3 – Domínio da interpretação

Segundo Gersting (2001), uma interpretação para uma expressão envolvendo predicados consiste em:

- Uma coleção de objetos (domínio da interpretação) que precisa incluir pelo menos um objeto;
- A especificação de uma propriedade dos objetos no domínio para cada predicado na expressão;
- A atribuição de um objeto particular no conjunto universo para cada símbolo constante na expressão.

Alguns exemplos auxiliam a tarefa de compreender a interpretação I :

$$\begin{array}{ll} (\Delta^I) = \{d_1, \dots, d_n, n > 1\} & (\Delta^I) = \{1, 2, 3, \dots\} \\ a^I = D_1 & 1^I = 1 \\ b^I = D_2 & 2^I = 2 \end{array}$$

A.3) Sobre as definições de intensional e extensional

Outros conceitos que devem ser introduzidos são as definições *intensionais* e definições *extensionais*. Segundo o *ISO-standard 704*, uma definição *intensional* consiste de uma lista de características do conceito a ser definido. Com esse propósito, são fornecidos o gênero mais próximo já definido (ou que se espera ser conhecido) e as características que determinam

esse gênero. Uma ou mais dessas características também diferenciam o conceito em relação a outros conceitos da mesma série horizontal. Por exemplo, lâmpada incandescente: lâmpada elétrica, na qual um filamento é aquecido por uma corrente elétrica de forma que a lâmpada emita luz. A lâmpada incandescente é definida como o auxílio do gênero mais próximo (lâmpada elétrica) e de suas características (filamento, luz emitida a partir do aquecimento por corrente elétrica). As características distinguem tais lâmpadas de todos os outros tipos de lâmpadas elétricas, as quais, junto aos conceitos mencionados acima, compõem as séries horizontais.

Uma definição *extensional* consiste de uma enumeração de aspectos de todas as espécies que são do mesmo nível de abstração. Algumas vezes, todos os objetos individuais podem ser enumerados. A definição *extensional* pode não ser válida por muito tempo, pois novas espécies podem aparecer no mesmo gênero. Em geral, uma definição *extensional* é mais facilmente compreendida que uma *intensional*. Por exemplo, “os planetas do sistema solar são Mercúrio, Vênus, Terra, Marte, Júpiter, Saturno, Urano, Netuno e Plutão”.

A figura abaixo apresenta mais um exemplo para melhor entendimento da idéia de definições *intensionais* e *extensionais*:

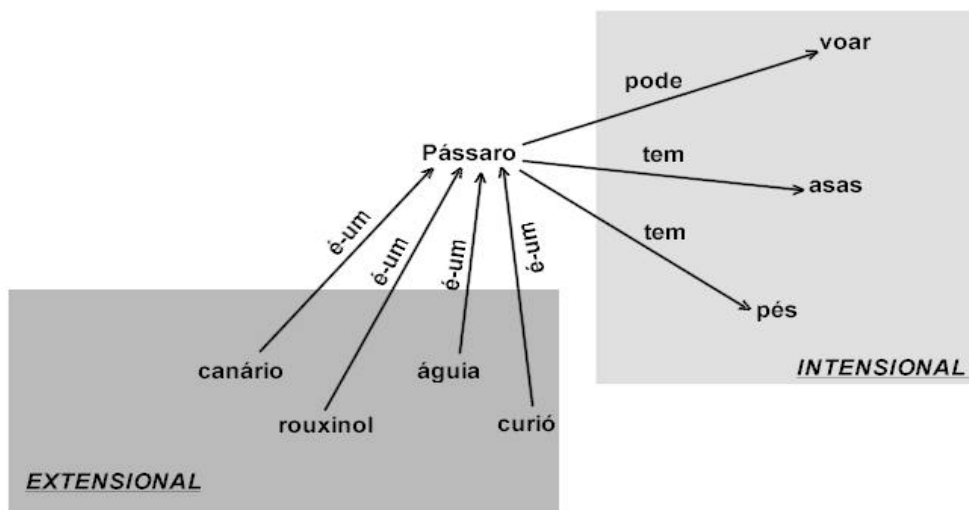


Figura A4 – Definições intensionais e extensionais

Para Calvanese et al. (1998), a definição *extensional* é aquela que expressa relacionamentos entre as extensões de conceitos envolvidos. A definição *intensional* é aquela que expressa relacionamentos conceituais que não necessariamente se refletem no nível das instâncias. No processo de integração as *assertivas inter-modelos* serão consideradas como *intensionais* e *extensionais*.

Uma característica dos sistemas baseados em Lógica Descritiva é que a base de

conhecimento é construída a partir de dois componentes. O primeiro, chamado *T-Box* (*terminological part*), é um esquema global que lida com as classes de indivíduos, suas propriedades gerais e relacionamentos. O segundo, chamado *A-Box* (*assertional part*), é uma instânciação (parcial) do esquema, que contém assertivas relacionando indivíduos a classes, ou um indivíduo a outro. Essa nomenclatura é proveniente da *KL-One* (BRANCHMAN; SCHMOLZE, 1985¹⁰⁷ citado por DONINI et al., 1997).

O *T-Box* corresponde ao grupo de axiomas que descrevem a estrutura do domínio e pode ter *axiomas de definição* e *axiomas de inclusão*. *Axiomas de definição* introduzem nomes para os conceitos, por exemplo (DECKER et al., 2000) (a simbologia usada nos exemplos será explicada em seguida):

$$Pai = Homem \sqcap E \text{ tem-criança.humana} \quad (II.1)$$

$$Homem \sqsubseteq animal \sqcap bípede \quad (II.2)$$

Na sentença (II.1), define-se que o pai (*expressão de conceito*) é um indivíduo homem (*conceito* ou *papel*) que tem uma criança (*cadeia de características*, que faz parte do *papel*), que é humana (predicado, o qual também faz parte do *papel*). Os símbolos $\sqcap$ e E , correspondem, respectivamente, ao construtor *interseção* e ao quantificador *existencial*. Baader et al. (2003), definem formalmente a *restrição de predicado existencial* na forma

$$(u_1, \dots, u_n).P$$

onde $u_1, \dots, u_n$ são características e P é um predicado de aridade n . Uma forma mais genérica é, por exemplo a expressão:

$$E \text{ (temIdade)(temIdade)}>.$$

que consiste de todos o par de indivíduos que tem uma idade de forma que o primeiro é sempre mais velho que o segundo. A formalização desta idéia também é apresentada por (BAADER et al., 2003):

$$E(u_1, \dots, u_n) (u_1, \dots, u_n).P$$

onde P é um predicado de aridade n e v e u são as cadeias de características.

¹⁰⁷ BRANCHMAN, R.; SCHMOLZE, J. An overview of the KL-One, Knowledge representation system *Cognitive Science*, v. 9, n. 2, p. 171-216, 1985.

Na sentença (II.2), define-se que homem (*expressão de conceito*) é um animal bípede (interseção das *expressões de conceito* animal e bípede).

Axiomas de inclusão especificam relações de propriedade entre relações, como por exemplo (DECKER et al., 2000):

$$E \text{ tem-graduação.mestre} \sqsubseteq E \text{ tem-grau.bacharel} \quad (\text{II.3})$$

Na sentença (II.3), ter a graduação de mestre inclui ter a graduação de bacharel. Ou seja, existe um *papel*, representado por relação binária composta pelos termos “*tem-graduação*” e “*mestre*”, já está “inclusa” no papel formado pelos termos “*tem-grau*” e “*bacharel*”. Para ser mestre, o indivíduo já tem que ser bacharel.

O *A-Box* é grupo de axiomas que descrevem situações concretas, sendo composto por *assertivas de conceitos* e *assertivas de papéis* (essas assertivas também são axiomas). Um exemplo para as *assertivas de conceitos* é apresentado por (DECKER et al., 2000):

$$\text{João} : \text{homem} \sqsupset E \text{ tem-crianças.fêmeas} \quad (\text{II.4})$$

Na sentença (II.4), João é um homem (conceito ou expressão de conceito) que tem crianças do sexo feminino (definido pelo *papel* “*temCrianças.fêmeas*”). Um exemplo para *assertivas de papéis* é:

$$\langle \text{João}, \text{Maria} \rangle : \text{tem-crianças} \quad (\text{II.5})$$

ou seja, João e Maria, simplesmente, têm crianças.

A *lógica terminológica* é um formalismo, baseado em *linguagens conceituais*, que permite a classificação, descrição de relações entre classes e, especialmente, a definição de uma estrutura hierárquica. As linguagens conceituais apresentam vantagens, se comparadas com outros formalismos para representação do conhecimento (BAADER et al., 1992):

- Possuem uma semântica declarativa bem definida, o que quer dizer que o significado da construção não é dado operacionalmente. Por exemplo, na afirmação “João é pai”, esse fato não ocorre simplesmente por que o sistema responde à pergunta “O que é João ?” com a afirmação “Pai”, mas em função da sua descrição e seu modelo. Assim “João é pai” em todos os modelos, em todas as situações que a descrição é usada;
- Existe uma caracterização de tarefas dos sistema de representação do conhecimento através da semântica declarativa;

- Existem vários algoritmos que realizam estas tarefas e suas propriedades já são bem conhecidas (por exemplo, completude, decidibilidade, complexidade, dentre outras).

A recuperação da informação em sistemas de lógica descritiva é um processo dedutivo que envolve o esquema (*T-Box*) e sua instanciação (*A-Box*). O *T-Box* não é apenas uma serie de restrições sobre o *A-Box*, pois contém informação intensional sobre as classes, a qual é considerada quando se respondem consultas na base de conhecimento.

A parte *T-Box* também é chamada de *conhecimento intensional* da base de conhecimento. Existem diversas propostas para expressar o conhecimento intensional e uma difere da outra pela diversas construções que são permitidas. Dada uma linguagem *L*, uma declaração intensional pode ter as seguintes formas (DONINI et al., 1997)¹⁰⁸:

$$A \sqsubset C \quad \text{Especificação de conceito primitivo} \quad (\text{II.6})$$

$$A = C \quad \text{Definição de conceito} \quad (\text{II.7})$$

$$C \sqsupset D \quad \text{Inclusão de conceitos} \quad (\text{II.8})$$

$$C = D \quad \text{Equivalência de conceitos} \quad (\text{II.9})$$

onde "*A*" é um *nome de conceito* e "*C*" e "*D*" são conceitos na linguagem *L*. É importante lembrar, conforme citado anteriormente, que, as letras "*A*" e "*B*" são usadas como *conceitos atômicos*, e, "*C*" e "*D*" para *expressões de conceito* (formadas por *construtores de conceitos*).

O que é chamado *primitive-T-Box*, consiste apenas de declarações do tipo (II.6) e tem boas propriedades computacionais, sendo por isso muito utilizado em *bancos de dados orientados à objeto* e *modelos de dados semânticos*. Uma "*simple-T-Box*" é formada por um grupo de declarações do tipo (II.6) e (II.7) e são os mecanismos mais utilizados em sistemas já implementados (por exemplo o *Classic*, apresentado e o *Back*, citados na Seção 4.3.5).

Na *simple-T-Box*, um *nome de conceito* que aparece no lado esquerdo da definição do conceito, é chamado de *conceito-definido*; um *nome de conceito* que aparece do lado esquerdo de um especificação de conceito é chamada *conceito-primitivo*; e, um *nome de conceito* que não aparece do lado esquerdo da declaração é chamado de *conceito-atômico*. Um tipo mais genérico é conhecido como *free-T-Box*, que tem a forma das sentenças (II.8) e (II.9). Neste tipo, a distinção entre os conceitos definidos, primitivos e conceitos atômicos não tem importância, pois as definições não envolvem *nomes de conceitos*.

¹⁰⁸ BAADER et al. (1992, p. 6, seção 5 *Terminological Axioms*) também descreve esses axiomas mas utilizando outros símbolos. O significado para as duas apresentações é o mesmo.

As regras definidas sobre a parte *intensional*, com o uso de *interpretações*, são utilizadas um pouco adiante para integração.

A parte *A-Box* também é chamada de *conhecimento extensional* da base de conhecimento. A construção do conhecimento *extensional* é feita permitindo-se que conceitos e papéis sejam usados em assertivas sobre indivíduos. Dada uma linguagem L , uma declaração *A-Box* em L tem as seguinte formas:

$$C(a) \quad \text{assertiva de associação de conceitos} \quad (II.10)$$

$$R(a,b) \quad \text{afirmativa de associação de papéis} \quad (II.11)$$

onde C é um conceito de L , R é um papel de L e “ a ” e “ b ” são indivíduos.

Também são definidas regras semânticas para a assertivas (II.10) e (II.11). Novamente através da noção de interpretação (explicada adiante) será possível obter integração.

Em resumo, dada uma *linguagem conceitual* L , uma *base de conhecimento* Σ é um par $\Sigma=(T, A)$, onde T é a parte *intensional* e A é a parte *extensional*. Essas duas partes, que definem uma base de conhecimento, serão consideradas ao ser estudar a integração semântica das fontes de dados. O entendimento das definições *intensionais* e *extensionais* é importante para a abordagem aqui utilizada.

Anexo B – O processo de integração baseado em lógica descritiva

Para que o meta-modelo possa ser representado formalmente em lógica descritiva, mais algumas considerações são necessárias. Em seguida a estas considerações, apresentam-se as condições em que as *assertivas inter-modelos* podem proporcionar integração. Para facilitar a leitura do formalismo, apresenta-se uma breve descrição do raciocínio utilizado para se chegar à integração:

- 7) Um modelo conceitual, que corresponde ao meta-modelo do domínio (MMD), consiste de *assertivas intra-modelo*, que expressam o conhecimento dos *conceitos* e *relações* do modelo, e, de *assertivas inter-modelo*, que expressam o conhecimento do relacionamento entre fontes distintas;
- 8) As *assertivas intra-modelo* são representadas formalmente pela sentença (14);
- 9) As *relações* e os *conceitos*, cujo conhecimento é descrito pelas *assertivas intra-modelo*, são definidos pelas sentenças (1) e (2);
- 10) Para que uma *interpretação* satisfaça a um conjunto de *assertivas intra-modelo*, as condições expressas nas sentenças (15), (16) e (17) devem ser atendidas;
- 11) No seu aspecto *intensional*¹⁰⁹, as *assertivas inter-modelo* devem atender a condições idênticas às *assertivas intra-modelo*, as quais serão apresentadas; no aspecto *extensional*¹¹⁰, a demonstração tem algumas diferenças e será omitida;
- 12) Obtém-se o meta-modelo ao se conseguir que todas as *assertivas intra-modelo* e *assertivas inter-modelo* sejam atendidas;
- 13) Utilizando serviços de inferência¹¹¹ sobre o meta-modelo, podem-se deduzir automaticamente conseqüências lógicas, mesmo que nem todas as relações envolvidas estejam presentes em todos os modelos.

B1) A representação em lógica descritiva

As lógicas descritivas¹¹² (também conhecidas como lógicas terminológicas) foram introduzidas no início dos anos 80 como uma tentativa de proporcionar bases formais às *redes*

¹⁰⁹ Relativas à parte terminológica da base de conhecimento (T-Box); maiores detalhes sobre esse tipo de definição são encontrados no Anexo B;

¹¹⁰ Relativas à parte assertiva da base de conhecimento (A-Box); maiores detalhes sobre esse tipo de definição são encontrados no Anexo B;

¹¹¹ Os serviços de inferência são brevemente discutidos ao final do Anexo C.

¹¹² Ver <<http://dl.kr.org>>.

semânticas. Um dos problemas das *redes semânticas* era justamente que sua semântica não era clara. Desde então as lógicas descritivas vêm evoluindo como *linguagens conceituais*, que podem abranger praticamente todos os formalismos de representação baseados em classes, utilizados em taxonomias, thesaurus, terminologias, esquemas de classificação e bancos de dados.

Um sistema de representação do conhecimento em lógica descritiva é concebido através de uma *linguagem conceitual* (ou linguagem terminológica) que consiste de um grupo de construções que representam *classes* e *relacionamentos* entre essas *classes* em um domínio do conhecimento. Nessas *linguagens conceituais*, *classes* são representadas por *conceitos* (os quais são usados para especificar um grupo de indivíduos) e *papéis* (os quais são relações binárias usadas para especificar propriedades ou atributos) (DONINI et al., 1997).

Segundo Baader et al. (1992), uma *linguagem conceitual* é baseada em *primitivas*, ou seja, axiomas terminológicos, os quais são símbolos de um alfabeto. As primitivas que formam a linguagem são:

- *Conceitos atômicos*, também chamados de *nomes de conceitos*¹¹³;
- *Papéis atômicos*, também chamados de *nomes de papéis*;
- *Nomes de atributos*;
- *Nomes de indivíduos*;
- *Nomes de objetos*.

Nomes de atributos são papéis de cardinalidade um. Em geral, o termo *papel* é utilizado quando um *relacionamento* é descrito no contexto de dois conceitos relacionados e o termo *atributo* é usado quando um *relacionamento* é descrito como uma propriedade de um *conceito*. *Nomes de Indivíduos* são *conceitos*, *papéis* e *instâncias*. O elemento básico do nível das instâncias é o *objeto*, que tem como identificador um símbolo único. De forma a refletir associações relevantes, os *objetos* podem ser agregados em *tuplas*. O número de objetos que são componentes de uma *tupla* é chamado de *aridade*. *Tuplas* de mesma *aridade* podem ser agrupadas em conjuntos chamados *relacionamentos*, os quais são descritos por um conjunto de *papéis*. Um *papel* representa um componente de um *relacionamento*, o qual tem tantos *papéis* quanto sua *aridade*.

Com essas primitivas é possível formar expressões mais complexas:

- *Expressões de conceitos*, também chamadas de *descrições de conceitos*;

¹¹³

As duas formas são encontradas na literatura e apresentadas aqui para maior clareza;

- *Expressões de papéis*, também chamadas de *descrições de papéis*;
- *Expressões de atributos*, também chamadas de *descrições de atributos*.

As expressões acima são obtidas pelo uso de *construtores* sobre as *primitivas*, definindo novos *conceitos* e *papéis*. Dessa forma, além dos *conceitos atômicos*, *papéis atômicos* e *nomes de atributos*, o alfabeto contém ainda *operadores para formação de conceitos*, *operadores para formação de papéis e atributos*, *axiomas terminológicos* e *axiomas assertivos* (para uma descrição da sintaxe, semântica e exemplo dos principais operadores ver Figura B1 e Figura B2 do Anexo B; para uma lista mais completa de construtores ver Baader et al. (1992)).

Um exemplo dos *operadores de formação de conceitos* é o conceito “mãe”, que pode ser descrito como:

<i>Mãe</i>	=	<i>Pessoa</i>	é	(<i>sexo</i>	:	<i>feminino</i>)
<i>Expressão de conceito</i>		<i>conceito atômico</i>		<i>nome de atributo</i>		<i>nome de indivíduo</i>

Um exemplo dos *operadores de formação de papéis e atributos* é o conceito “filha”, que pode ser descrito como:

<i>Filha</i>	=	<i>parente_feminino</i>	é	<i>criança</i>
<i>Expressão de papel</i>		<i>papel</i>		<i>papel</i>

Os *axiomas terminológicos* (definições, especializações e restrições) são usados para especificar o conhecimento sobre o mundo. Um grupo de *axiomas terminológicos* determina uma *terminologia*. Além disso, tem-se os *axiomas assertivos*, que preenchem o mundo com objetos.

Todos esses elementos são componentes de uma linguagem formal utilizada para projetar uma *base de conhecimento*.

A lógica descritiva pode especificar os componentes de um sistema de integração de dados, conforme será visto adiante. O formalismo utilizado para tal é baseado em lógica descritiva *DLR*¹¹⁴, cujos elementos básicos são os relações unárias (também chamadas *conceitos*) e as *relações “n-árias”* (chamadas “de aridade¹¹⁵ n”).

Em *DLR*, por convenção, a letra “A” é usada para *conceitos atômicos*, e, “C” para

¹¹⁴ A sigla “DLR” representa um dos possíveis dialetos em uso para a lógica descritiva, que apresenta extensões desenvolvidas por um pesquisador ou comunidade. Cada dialeto define “construtores” (conjunção, disjunção, etc.) que podem ser usados recursivamente para definir novos conceitos e papéis.

¹¹⁵ Número de papéis associados a uma relação, por exemplo: a relação “casal” tem aridade 2, ou seja, marido e mulher.

expressões de conceitos. Ao se trabalhar com um alfabeto¹¹⁶ Z , as sentenças (1) e (2) abaixo descrevem, em BNF-Backus-Naur Form¹¹⁷, a gramática da linguagem de expressão da DLR. Os conceitos arbitrários C (expressões de conceitos) e as relações arbitrárias R (de aridade entre 2 e n) são assim construídos pela sintaxe das sentenças (1) e (2), e:

$$R ::= T_n \mid P \mid (\$/n:C) \mid ! R \mid R_1 \sqcap R_2 \quad (1)$$

onde, na sentença (1):

- “ n ” é a aridade da relação;
- “ T_n ” é uma *primitiva*, subconjunto do domínio interpretação (este último representado na sentença (2) por T_1), que cobre todas as relações de aridade n ;
- “ P ” é uma *primitiva*, que representa um grupo finito de *relações atômicas*;
- “ $(\$/n:C)$ ” representa a *relação* de aridade n , que consiste de tuplas nas quais o i -ésimo componente pertence a C ; i e n são componentes das relações;
- “ $! R$ ” significa negação de R , seu complemento, todas as relações que sobram sem o grupo R ;
- $R_1 \sqcap R_2$ representa a conjunção entre duas relações R_1 e R_2 .

Traduzida para o português, a sentença (1) pode ser escrita:

As relações arbitrárias R consistem da primitiva T_n ou um grupo finito de relacionamentos atômicos ou relacionamentos de aridade n ou o complemento das relações arbitrárias R em relação ao domínio da interpretação ou a conjunção de duas relações arbitrárias.

$$C ::= T_1 \mid A \mid ! C \mid C_1 \sqcap C_2 \mid E [\$i] R \mid () K [\$i] R \quad (2)$$

e na sentença (2):

- “ T_1 ” é uma primitiva que representa o domínio da interpretação;
- “ A ” é uma primitiva que representa um grupo de conceitos atômicos;
- “ $! C$ ” significa negação de C , seu complemento, todas as conceitos que sobram sem o grupo C ;
- “ $C_1 \sqcap C_2$ ” representa a conjunção entre dois conceitos arbitrários C_1 e C_2 , quaisquer;

¹¹⁶ Um alfabeto consiste de uma lista de símbolos dividida em entidades, relacionamentos, papéis, atributos e valores.

¹¹⁷ Uma meta-sintaxe formal usada para expressar gramáticas.

- “E [\$i] R”; se R é um relacionamento de aridade n, sendo $i \in \{1, \dots, n\}$, então a expressão representa o conceito que agrupa os indivíduos que ocorrem como i-ésimo componente de alguma tupla de R;
- “($\bigvee$ K [\$i]R)” representa os relacionamentos R que atendem à restrição K (número inteiro não negativo), ou melhor, representa o conceito que agrupa aqueles indivíduos “d” para os quais o grupo de relacionamentos R contém no máximo K tuplas, das quais o i-ésimo componente é “d”;

Traduzida para o português, a sentença (2) pode ser escrita:

O conjunto de conceitos arbitrários C consiste da primitiva T_1 , a qual representa o domínio da interpretação ou um grupo de conceitos atômicos ou o complemento ao próprio conjunto de conceitos arbitrários ou a conjunção de dois conjuntos de conceitos arbitrários ou o conceito que agrupa os indivíduos que ocorrem como i-ésimo componente de alguma tupla de R ou o conceito que agrupa aqueles indivíduos “d” para os quais o grupo de relacionamentos R contém no máximo K tuplas, das quais o i-ésimo componente é “d”;

É importante ainda considerar que apenas relações de mesma aridade podem ser combinadas para formar expressões do tipo $R_1 \sqcap R_2$ (a qual herda a aridade n), e que, $i \in n$, sempre que i representa um componente da relação de aridade n.

B2) O formalismo de integração de fontes

A atividade de resolução de conflitos semânticos pode ser executada com base em técnicas de integração de bancos de dados, visto que os problemas são análogos àqueles problemas estruturais e semânticos que existem em um modelo de dados *entidade-relacionamento* (CALVANESE et al., 1998). Um exemplo é a situação em que o mesmo conceito é representado como uma classe em um modelo e como uma relação em outro modelo. Esses princípios estão bem estudados e estabelecidos em Batini, Ceri e Navathe (1992)¹¹⁸ citados por Calvanese et al. (1998).

As construções semânticas da lógica DLR são especificadas através da noção de *interpretação*¹¹⁹. Uma *interpretação* $I = (\Delta^I, \cdot^I)$ é constituída por um domínio de interpretação Δ^I e uma função de interpretação $\cdot^I$ que atribui a cada conceito C um subconjunto C^I de Δ^I e, a cada

¹¹⁸ BATINI, C.; CERI, S.; NAVATHE, S. B. Conceptual database design, an entity-relationship approach. & Navathe, S. 1992. Conceptual Database Design: An Entity-Relationship Approach. Redwood City, California Addison Wesley Higher Education Copyright, 1992. 470 p.

¹¹⁹ Maiores detalhes sobre a interpretação são fornecidos no Anexo B, Seção B2.

relação R de aridade n atribui um subconjunto R^I de Δ^I , de forma que regras semânticas da DLR sejam atendidas. Exemplos da semântica para uma interpretação $I = (\Delta^I, .^I)$ são apresentados nas Figuras B1 e B2 do Anexo B (NEBEL, 1998). As regras semânticas que devem ser atendidas são as seguintes:

$$T_n^I \subseteq (\Delta^I)^n \quad (3)$$

$$P^I \subseteq T_n^I \quad (4)$$

$$(!R)^I = T_n^I / R^I \quad (5)$$

$$(R_1 \sqcap R_2)^I = R_1^I \cap R_2^I \quad (6)$$

$$(\$i / n:C)^I = \{(d_1, \dots, d_n) \in T_n^I \mid d_i \in C^I\} \quad (7)$$

$$T_1^I = \Delta^I \quad (8)$$

$$A^I \zeta \Delta^I \quad (9)$$

$$(!C)^I = \Delta^I / C^I \quad (10)$$

$$(C_1 \sqcap C_2)^I = C_1^I \cap C_2^I \quad (11)$$

$$(E[\$i]R)^I = \{d \in \Delta^I \mid \exists \{(d_1, \dots, d_n) \in R^I . d_i = d\}\} \quad (12)$$

$$(\bigvee_k [\$i]R)^I = \{d \in \Delta^I \mid \exists \{(d_1, \dots, d_n) \in R^I . d_i = d \mid \bigvee_k\} \} \quad (13)$$

O modelo conceitual MMD corresponde a um conjunto finito de *assertivas intra-modelo* que expressam o conhecimento das relações e conceitos de MMD e tem a seguinte forma:

$$L_1 \sqsubseteq L_2 \quad L_1 \not\sqsubseteq L_2 \quad L_1 \equiv L_2 \quad L_1 \neq L_2 \quad (14)$$

onde L_1 e L_2 são duas expressões de classe (relações ou conceitos) de mesma aridade, sobre o alfabeto Z do esquema S ($S = \langle Z, \Sigma \rangle$ esta definido no segundo parágrafo adiante).

A interpretação I atende aos requisitos da “assertiva intra-modelo” se:

$$L_1 \sqsubseteq L_2 \quad (15)$$

se for atendida a condição

$$L_1^I \subseteq L_2^I \quad (16)$$

e, atende a “assertiva intra-modelo”:

$$L_1 \not\sqsubseteq L_2 \quad (17)$$

se não satisfaz a primeira afirmativa.

Uma interpretação I é chamada modelo de um grupo de assertivas Σ se cada assertiva do grupo é satisfeita por I . O modelo possui um esquema S representado por $S = \langle Z, \Sigma \rangle$ onde Z é o alfabeto de símbolos e Σ é um grupo de *assertivas intra-modelo* (ou intra-esquema) sobre Z . A semântica de S é formalmente caracterizada por todas as interpretações para S que satisfaçam cada assertiva em Σ , ou seja, todos os modelos de S . Para expressar o conhecimento das inter-relações conceituais entre as fontes e/ou entre o modelo genérico, usam-se as *assertivas inter-modelo*. Para satisfazê-las, as condições são análogas àquelas para as *assertivas intra-modelos*.

Dada uma linguagem de conceitos L , uma base de conhecimento Z em L é um par $Z = (T, A)$ onde T é a parte *intensional* (*T-Box*) e A a parte *extensional* (*A-Box*)¹²⁰. Uma interpretação I é um modelo para Z , se e somente se, é um modelo para T e um modelo para A .

As assertivas inter-modelos podem ser extensionais ou intensionais. A interpretação das assertivas inter-modelo extensionais é análoga às condições para as assertivas intra-modelos apresentadas acima. Já as assertivas inter-modelo intensionais têm uma interpretação diferente, a qual será omitida aqui.

Se MMD é um *meta-modelo de domínio* em um conjunto de $(m+2)$ elementos

$$\langle M_0, M_1, \dots, M_m, G \rangle$$

tal que:

- M_0 é o MG (ou seja, o modelo genérico);
- Cada M_i , para $i \in \{1, \dots, m\}$ é um MF (ou, seja um modelo fonte);
- G é um conjunto finito de *assertivas inter-modelo* que devem ser satisfeitas;

uma interpretação I satisfaz a MMD se satisfaz todas as *assertivas intra-modelos* e *inter-modelos* em MMD.

Como uma base de conhecimento em lógica descritiva consiste de um grupo de axiomas terminológicos intensionais (*T-Box*) e um grupo de axiomas assertivos extensionais (*A-Box*), se as condições (axiomáticas) são satisfeitas por uma interpretação, então se pode dizer que essa interpretação é um modelo dos axiomas.

¹²⁰ Maiores informações sobre as definições intensionais e extensionais são apresentadas no Anexo B, Seção B3.

B3) Exemplo de uso de assertivas inter-modelo para integração

As sub-relações apresentadas nesse exemplo podem ser obtidas diretamente das regras especificadas para os conceitos e relações no meta-modelo. O modelo *entidade-relacionamento* utilizado na modelagem de bancos de dados é totalmente compatível com a lógica descritiva (CALVANESE et al., 1998). Apresentam-se a seguir as sub-relações e suas dependências, representando as *assertivas intra-modelos*. Também são apresentadas as *assertivas inter-modelo* que atendem às regras semânticas (especificadas na seção 5.4.3). Os modelos são integrados, no nível conceitual, sendo que o resultado é uma ontologia de mais alto nível que combina os outros modelos.

Seja MF₂, o modelo-fonte e as *assertivas intra-modelos* seguintes:

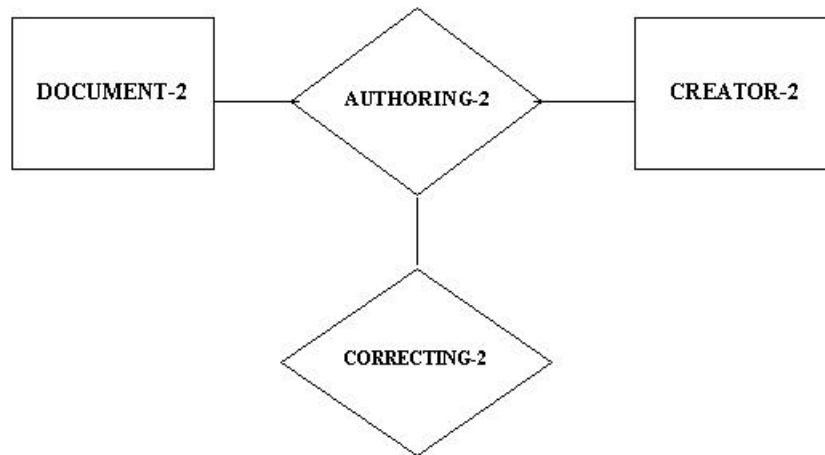


Figura B1 – Modelo para as assertiva intra-modelo MF₂

$$\text{Authoring2} \sqsubseteq (\$1:\text{Document-2}) \sqcap (\$2:\text{Creator-2})$$

$$\text{Correcting-2} \sqsubseteq \text{Authoring-2}$$

Seja MF₃, o modelo-fonte e as *assertivas intra-modelos* seguintes¹²¹:

¹²¹ Considera-se, conforme restrições da classe document”, que um creator0 pode publicar (por exemplo, na Internet).

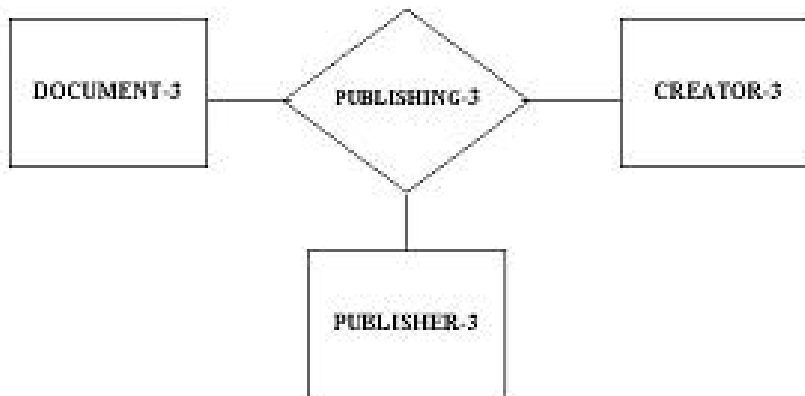


Figura B2 – Modelo para a assertiva intra-modelo MF3

$\text{Publishing-3} \sqsubseteq (\$1:\text{Document-3}) \lceil (\$2:\text{Creator-3}) \lceil (\$3:\text{Publisher-3})$

Seja MMD (representado pelo índice zero) o meta-modelo, ou seja, a visão reconciliada:

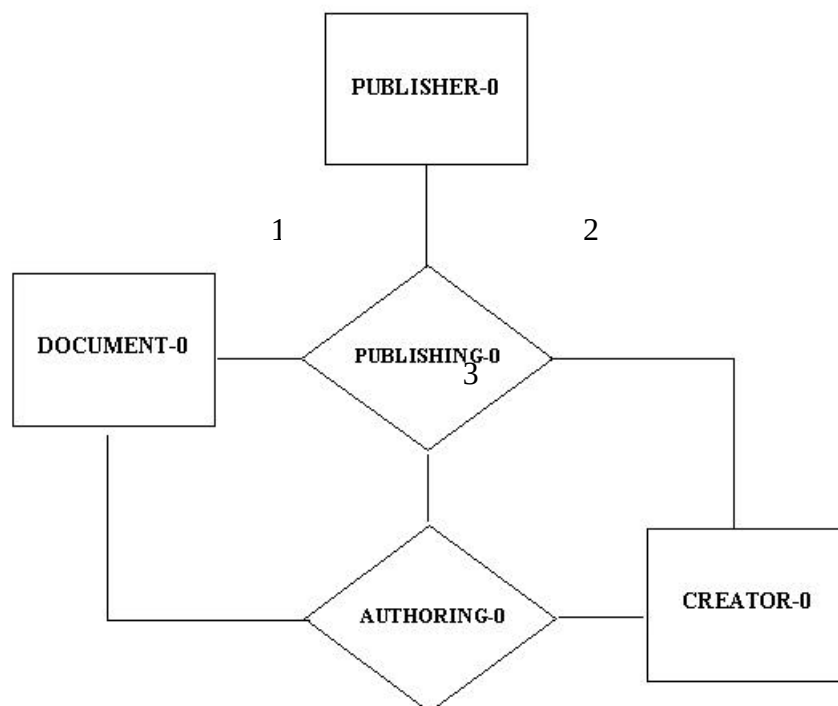


Figura B3 – Visão reconciliada

A partir das regras semânticas determinadas neste anexo, seção B3, as sentenças abaixo representam as *assertivas inter-modelos* e dizem respeito às várias interdependências entre as relações e conceitos no modelo-genérico e os dois modelos-fonte, conforme segue¹²²:

¹²² Nas sentenças “*int*” quer dizer intensional e “*ext*” quer dizer extensional.

$$\begin{aligned}
& \text{Authoring-2} \sqsubseteq_{\text{ext}} \text{Authoring-0} \\
& \text{Document-2} \uparrow_{\text{ext}} \text{Document-0} \sqcap \text{E} \sqcap \text{Authoring-0} [\$1].T_2 \\
& \text{Document-0} \sqcap \text{E} \sqcap \text{Publishing-0} [\$1].T_3 \sqsubseteq \text{E} \sqcap \text{Authoring-0} [\$1].T_3 \\
& \text{Document-2} \sqsubseteq_{\text{ext}} \text{Document-0} \sqcap \text{E} \sqcap \text{Publishing} [\$1].T_3 \\
& \text{Document-3} \uparrow_{\text{ext}} \text{Document-0} \\
& \text{Publisher-3} \uparrow_{\text{ext}} \text{Publisher-0} \\
& \text{Document-2} \uparrow_{\text{int}} \text{Document-3} \\
& \text{Creator-2} \uparrow_{\text{int}} \text{Creator-3}
\end{aligned}$$

B4) Sobre serviços de inferência

O conceito de inferência é obtido na lógica, especificamente na análise de argumentos demonstrativos. Pinto (2001, p.19/20), define inferência como uma relação de caráter imaterial, que conduz de um antecedente a um conseqüente. A figura a seguir apresenta um exemplo:

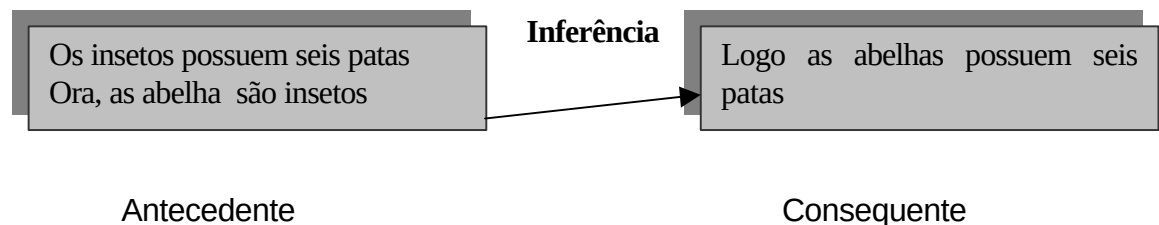


Figura B4 – Ilustração do significado de “inferência”

Para Baader et al. (1992) os sistemas de representação do conhecimento devem oferecer serviços que possibilitem a organização, o gerenciamento, a modificação ou a recuperação da informação distribuída hierarquicamente. Estes são os serviços de inferência (ou de dedução). O sistema deve ser capaz, a partir do uso desse serviços, de responder às seguintes questões:

- Um conceito introduzido é definido de maneira significativa para todos? (satisfabilidade)
- Um conceito é mais geral que o outro? (hierarquia)
- Onde exatamente o conceito está situado na hierarquia de conceitos? (classificação)

- A representação do conhecimento é consistente? (consistência)
- Quais fatos são dedutíveis a partir do conhecimento? (instanciação)
- De quais conceitos um objeto é instância? (significação)
- Quais são as instâncias de um determinado conceito? (recuperação)

No exemplo apresentado na seção anterior, pode-se automaticamente derivar conseqüências lógicas do meta-modelo, a partir deste tipo de serviço, associado à DLR. Por exemplo, pode-se afirmar que a assertiva apresentada abaixo é uma conseqüência lógica do meta-modelo:

$$\text{Correcting-2} \sqsubseteq_{\text{ext}} \text{Authoring-0} \sqsupset \text{Creator-0}$$

É importante observar que embora o modelo-genérico (MG) não contenha a relação “*correcting*”, a assertiva acima relaciona “*correcting-2*” de forma precisa.

Este trabalho não se aprofunda no assunto de inferências. Outras técnicas são apresentadas por Catarci e Lenzerine (1993), Donini et al. (1997) e Patel-Schneider e Swarout (1993), que tratam de inferências em lógica descritiva. Livros de Inteligência Artificial também tratam desse assunto com mais detalhes.

Anexo C – Sobre o OIL, RDFe o RDFS

C1) OIL

O *Ontology Interchange Language (OIL)* (ou, para alguns autores *Ontology Inference Layer*) é uma linguagem para modelagem de primitivas em sistemas “*frame-based*”, os quais se baseiam na noção de conceitos e na definição de superclasses e atributos. Além disso são utilizadas em sistemas baseados em lógica descritiva.

O OIL é uma linguagem intuitiva para representação de ontologias, de semântica formal bem estabelecida e que tem similaridades com as linguagens para a Web (RDF, RDFS e XML) (FENSEL et al., 2001). Além disso, o OIL possibilita o uso de inferências que asseguram sua completude, correção e eficiência

Horrocks et al. (2001) definem as funções das camadas do OIL, com relação ao tipo de dados que tratam:

- No nível dos objetos, são descritas as instâncias da ontologia (este nível não foi abordado nesse trabalho);
- No primeiro nível de metadados, são fornecidas as definições da ontologia e a terminologia que é utilizada pelas instâncias; essa camada é conhecida como *ontology-definition*;
- O segundo nível de metadados descreve características da ontologia, tais como autor, nome, assunto, etc; esse nível é conhecido como *ontology-container*;

O OIL possui características em comum com RDFS. As camadas necessárias para definições semânticas, que utilizam OIL, RDF e RDFS (BROEKSTRA et al., 2001) são:

- Lógica: cuida da semântica formal e a inferências (usar o OIL);
- Esquema: trata da definição de vocabulário (usar o RDFS);
- Dados: contêm o modelo de dados e sintaxe de metadados (usar o RDF).

Na linguagem OIL, constrói-se uma ontologia através de um grupo de definições. A sintaxe compreende definições de classes (por exemplo, *class-def*), expressões de classes (construídas através de operadores *booleanos*), restrições (por exemplo, *slot-constraints*), dentre diversas outras¹²³.

¹²³ A sintaxe da linguagem OIL não será apresentada pois foge ao escopo do trabalho, mas pode ser obtida em www.ontoknowledge.com/oil.

C2) RDF/RDFS

O intercâmbio entre aplicações requer uniformidade sobre semântica, sintaxe e estrutura. Comunidades definem diferentes semânticas para seus metadados, de acordo com suas necessidades. A sintaxe, que consiste do arranjo sistemático de elementos de dados, facilita o intercâmbio e o uso de metadados entre diversas aplicações. Sobre a estrutura, pode-se imaginar uma limitação formal sobre a sintaxe, que proporcione uma representação consistente da semântica. Os metadados, ou são parte do próprio dado, ou estão separados dele, ou podem ser dados sobre outros metadados, configurando o que se chama de meta-meta-dados.

Dentre as novas propostas (a partir do uso intensivo da Web) está o RDF, uma infra-estrutura que possibilita a codificação, o intercâmbio e a reutilização de metadados, inspirada na especificação PICS. Trata-se de uma aplicação XML que impõe limitações estruturais à expressão semântica. Possibilita a publicação de vocabulários legíveis para pessoas e computadores, além de proporcionar a reutilização e a distribuição dos metadados .

Como infra-estrutura, o RDF permite a inter-operabilidade de metadados e permite a criação de mecanismos para convenções comuns de semântica, sintaxe e estrutura. O RDF utiliza o XML como sintaxe o que proporciona independência de plataforma, é legível para pessoas e possibilita representação de estruturas complexas.

O modelo de dados RDF proporciona uma forma para a descrição de “recursos”, os quais têm “propriedades” (atributos ou características). No contexto do RDF, um “recurso” é qualquer objeto identificável de forma única por uma URI (pode ser um conjunto de páginas da Web). As propriedades associadas aos recursos são chamadas “propriedades-tipo” e expressam os relacionamentos dos “valores” associados aos recursos. Os valores podem ser “atômicos” (*strings*, números, etc) ou outros recursos, os quais, por sua vez, podem ter suas propriedades. Às propriedades do recurso chama-se “descrição”.

Com a crescente necessidade de metadados para integração digital, o RDF não pode atender mais a todas expectativas em relação à semântica. Segundo DODDS et al.(2001), o RDF em conjunto com XML será suficiente para obter a representação desejada em muitos tipos de aplicações, mas , em alguns casos, será necessário maior poder de expressão.

Para estender o uso do RDF em situações que demandam maior número de metadados para a identificação dos dados, desenvolveu-se o RDFS (*RDF Schema*). A especificação RDFS é escrita em RDF e permite modelagem de um maior número de primitivas, proporciona informações sobre a interpretação das declarações do modelo RDF, permite verificar a integridade dos dados e estabelecer uma hierarquia de tipos, tais como classes, subclasses,

relacionamentos, restrições, dentre outras. Pode-se necessitar de mais informações sobre um recurso, “como por exemplo, o que ele representa: no caso de uma propriedade que representa um autor, pode ser necessário que o valor associado faça referência a uma pessoa, e não a um carro ou a uma casa; se uma propriedade representa um aniversário, é preciso verificar que o valor é uma data, e não um número ou um tipo de animal” (DODDS et al., 2001, p. 167).

O predicado de uma afirmação é identificado por uma URI e os objetos por *strings*. Além de efetuar a validação, o RDFS pode aplicar certas restrições ao uso de propriedades. O RDFS possibilita verificar se fazem sentido as propriedades aplicadas a um recurso e os valores atribuídos a estas propriedades.

Para maiores detalhes, consultar a especificação completa do RDFS, que contém todos os conceitos, expressões e operadores válidos¹²⁴. O esquema do RDFS definido em RDF também pode ser consultado¹²⁵.

¹²⁴ Disponível em: <<http://www.w3.org/TR/2000/CR-rdf-schema-20000327>>.

¹²⁵ Disponível em: <<http://www.w3.org/2000/01/rdf-schema>>.

Anexo D – Sintaxe do meta-modelo em OIL e RDFS

D1) Código em OIL

Na sintaxe abaixo estão descritas as relações entre conceitos do meta-modelo em OIL. Considera-se que MG é o modelo-genérico e MF1, MF2, ... MF_n são os modelos-fonte. A sintaxe em OIL apresentada abaixo foi gerada pelo OILED, diretamente a partir da definição de classes, propriedades, atributos, axiomas, etc, da Seção 5.4.2. Apresenta-se apenas um fragmento do código, em função do seu tamanho, mas que mostra exemplos das classes com seus respectivos *namespaces*.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:oiled="http://img.cs.man.ac.uk/oil/oiled#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#">
  <daml:Ontology rdf:about="">
    <dc:title>MEMO</dc:title>
    <dc:date>nov /2002</dc:date>
    <dc:creator>Mauricio Almeida</dc:creator>
    <dc:description>Trabalho para dissertação de mestrado ECI - UFMG
      - 2002</dc:description>
    <dc:subject>Meta-modelo para uso em projetos de integração de
      fontes de dados</dc:subject>
    <daml:versionInfo>1.0</daml:versionInfo>
  </daml:Ontology>
  <daml:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#format">
    <rdfs:label>format</rdfs:label>
    <rdfs:comment><![CDATA[]]></rdfs:comment>
    <oiled:creationDate><![CDATA[2002-11-29T23:04:34Z]]></oiled:creationDate>
    <oiled:creator><![CDATA[mba]]></oiled:creator>
    <daml:sameClassAs>
      <daml:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#format"/>
    </daml:sameClassAs>
  </daml:Class>
  <daml:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#creator">
    <rdfs:label>creator</rdfs:label>
    <rdfs:comment><![CDATA[]]></rdfs:comment>
    <oiled:creationDate><![CDATA[2002-11-29T22:46:48Z]]></oiled:creationDate>
    <oiled:creator><![CDATA[mba]]></oiled:creator>
    <rdfs:subClassOf>
      <daml:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#creator"/>
    </rdfs:subClassOf>
  </daml:Class>
  <daml:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#language">
    <rdfs:label>language</rdfs:label>
    <rdfs:comment><![CDATA[]]></rdfs:comment>
    <oiled:creationDate><![CDATA[2002-11-29T23:10:02Z]]></oiled:creationDate>
    <oiled:creator><![CDATA[mba]]></oiled:creator>
    <daml:sameClassAs>
      <daml:Restriction>
        <daml:onProperty rdf:resource="http://www.eci.ufmg.br/sabio#is-a"/>
        <daml:hasClass>
          <daml:Class>
            <daml:oneOf>
              <daml:List>
                <daml:first>
                  <daml:Thing
rdf:about="http://www.eci.ufmg.br/sabio#Portuguese"/>
                </daml:first>
                <daml:rest>
                  <daml:nil/>
                </daml:rest>
              </daml:List>
            </daml:oneOf>
          </daml:Class>
        </daml:hasClass>
      </daml:Restriction>
    </daml:sameClassAs>
  </daml:Class>
```



```

                                <daml:Restriction>
                                <daml:onProperty
rdf:resource="http://www.eci.ufmg.br/sabio#publishing"/>
                                <daml:hasClass>
                                <daml:Class
rdf:about="http://www.eci.ufmg.br/sabio/MG#document"/>
                                </daml:hasClass>
                                </daml:Restriction>
                                </daml:first>
                                <daml:rest>
                                <daml:nil/>
                                </daml:rest>
                                </daml:List>
                                </daml:rest>
                                </daml:List>
                                </daml:rest>
                                </daml:List>
                                </daml:rest>
                                </daml:List>
                                </daml:intersectionOf>
                                </rdfs:Class>
                                </daml:sameClassAs>
                                </daml:Class>

```

—————Código interrompido—————

D2) Código em RDFS

As relações entre conceitos do modelo em RDFS, correspondentes ao OIL são apresentados em seguida. A sintaxe em OIL abaixo foi gerada pelo OILed, diretamente a partir da definição de classes, propriedades, atributos, axiomas, etc, da Seção 5.3.3.

```

<?xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE rdf:RDF [
    <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
    <!ENTITY a 'http://www.w3.org/2000/01/rdf-schema#'>
]>
<rdf:RDF xmlns:rdf="&rdf;"
    xmlns:a="&a;">
    <rdf:Property rdf:about="file:C:/WINDOWS/Desktop/xx/instala_OILed/OilEd3.4FaCT/ontologies/memo2.daml#has-domain"/>
    <a:Class rdf:about="http://www.eci.ufmg.br/sabio#_MMD"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#authoring"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#correcting"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#has-one"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#is-a"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#is-published"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#publish-one"/>
    <rdf:Property rdf:about="http://www.eci.ufmg.br/sabio#publishing"/>
    <a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#creator">
        <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#creator"/>
    </a:Class>
    <a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#date">
        <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#date"/>
    </a:Class>
    <a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#document">
        <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#document"/>
    </a:Class>
    <a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#format">
        <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#format"/>
    </a:Class>

```

```

<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#language">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#language"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#publisher">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#publisher"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#subject">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#subject"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#title">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#title"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MF1#type">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#type"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#creator">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#creator"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#date">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#date"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#document">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#document"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#format">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#format"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#language">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#language"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#publisher">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#publisher"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#subject">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#subject"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#title">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#title"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MFn#type">
  <a:subClassOf rdf:resource="http://www.eci.ufmg.br/sabio/MG#type"/>
</a:Class>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#creator"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#date"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#document"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#format"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#language"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#publisher"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#subject"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#title"/>
<a:Class rdf:about="http://www.eci.ufmg.br/sabio/MG#type"/>
</rdf:RDF>

```

Anexo E – Metodologias, ferramentas, linguagens para construção e projetos que fazem uso de ontologias

E1) Metodologias

Metodologia	Breve descrição
<i>Cyc</i>	Codifica manualmente o conhecimento das fontes e, quando já tem conhecimento suficiente na ontologia, um novo consenso é obtido por ferramentas que utilizam linguagem natural. <i>Cyc</i> (LENAT; GUHA, 1990).
<i>Unschold</i>	Identifica o propósito, os conceitos e relacionamentos entre esses conceitos, além dos termos utilizados para codificar a ontologia e, em seguida, documentá-la. (USCHOLD; KING, 1996).
<i>Gruninger</i>	Método formal que identifica possíveis cenários de uso da ontologia, utiliza questões em linguagem natural para determinação do escopo da ontologia, executa a extração sobre os principais conceitos, propriedades, relações e axiomas, os quais são formalmente definidos em PROLOG ¹²⁶ (GRÜNINGER; FOX, 1995).
<i>KACTUS</i>	Método aplicado (recursivamente) que consiste em se fazer uma proposta inicial de uma base de conhecimento; em seguida, quando é necessária uma nova base em um domínio similar, propõe-se generalizar a primeira base em uma ontologia e adaptá-la a ambas as aplicações; quanto mais aplicações são construídas, mais genérica se torna à ontologia. <i>KACTUS</i> (BERNARAS; LARESGOITI; CORERA, 1996 ¹²⁷ citado por GOMEZ-PEREZ, (1999).
<i>Methontology</i>	A proposta é construir uma ontologia por reengenharia sobre outra ontologia; a construção é feita no nível do conhecimento relativo ao domínio e as atividades principais são: especificação, conceitualização, formalização, implementação e manutenção (FERNÁNDEZ-LÓPEZ et al., 1999).
<i>Sensus</i>	Baseia-se em abordagem <i>top-down</i> e deriva ontologias a partir de outras maiores; identificam-se termos relevantes para um domínio, os quais são conectados à ontologia mais abrangente (a <i>Sensus</i> , com cerca de 50.000 conceitos); os termos relevantes são selecionados automaticamente e um algoritmo monta a estrutura hierárquica do domínio. <i>Sensus</i> (SWARTOUT et al., 1996).
<i>On-to-knowledge</i>	Utilizada em projetos de gestão do conhecimento, auxilia a administração dos conceitos em organizações; identifica metas para as ferramentas de gestão do conhecimento e se baseia na análise de cenários e na atuação dos diversos provedores / clientes de informação da organização. <i>On-to-knowledge</i> (STAAB et al., 2001).

Figura E1 – Metodologias para construção de ontologias sem modelos, reutilização ou importação de outras ontologias

Metodologia	Breve descrição
CO4	Permite que usuários discutam sobre o conhecimento introduzido nas bases de conhecimento do sistema, as quais são compartilhadas; essas bases podem ser consideradas ontologias, pois deve haver um consenso sobre o conhecimento ali representado. CO4 (EUZENAT, 1996).
(KA) ²	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 coordenação são feitas via agentes inteligentes (KIETZ; MAEDCHE; VOLZ, 2000).

Figura E2 – Metodologias para construção de ontologias em grupo

Metodologia	Breve descrição
<i>Maedche</i>	É selecionada uma ontologia genérica, a qual será convertida em modelo. Na sequência especificam-se textos e obtém-se conceitos do domínio a partir das fontes disponíveis. São então removidos os conceitos genéricos da ontologia, de forma que apenas os conceitos específicos do domínio permaneçam. A estrutura conceitual da ontologia esta estabelecida e, baseada nela, obtém-se relações não taxonômicas dos textos. Novas relações conceituais são induzidas nos textos a partir da aplicação de métodos de aprendizado (KIETZ; MAEDCHE; VOLZ, (2000).

Figura E3 – Metodologias para aprendizado da estrutura de ontologias

¹²⁶ Linguagem de programação lógica implementada na França em 1972, que se tornou uma das mais utilizadas em inteligência artificial

¹²⁷ BERNARAS, A.; LARESGOITI, I.; CORERA., J. Building and Reusing Ontologies for Electrical Network Applications. *Proceedings of the European Conference on Artificial Intelligence, ECAI/96*, p. 298-302.

Metodologia	Breve descrição
<i>FCA-Merge</i>	Técnicas de processamento de linguagem natural e análise formal de conceitos são aplicadas sobre as ontologias que se deseja integrar, resultando em um conjunto de conceitos. O resultado é analisado por pessoas e transformado em uma ontologia composta. <i>FCA-Merge</i> (STUMME; MAEDCHE, 2001).
<i>PROMPT</i>	Baseado em um modelo de conhecimento genérico, trata-se de um algoritmo aplicável a várias plataformas integra ontologias. Executa algumas tarefas automaticamente (existe participação humana) e determina inconsistências na ontologia, sugerindo como removê-las. <i>PROMPT</i> (NOY; MUSEN, 2000).

Figura E4 – Metodologias para integração de ontologias

Metodologia	Breve descrição
<i>Guarino</i>	É uma metodologia para avaliação de taxonomias que utiliza princípios filosóficos. O usuário pode fazer anotações nas propriedades da taxonomia e depois avaliá-la (WELTY; GUARINO, 2000).
<i>Gomez-Perez</i>	Avalia erros na construção da ontologia, provenientes da estruturação do domínio em taxonomias e bases de conhecimento. Leva em conta o trabalho realizado previamente na avaliação das ontologias e os critérios utilizados para tal (GÓMEZ-PÉREZ, (2001).

Figura E5 – Métodos para avaliação de ontologias

Como métodos adicionais, que abordam pelo menos um aspecto da construção de ontologias pode-se citar: Ontolingua¹²⁸, CommomKADS¹²⁹ e Kactus¹³⁰, Plinius¹³¹, Onions¹³², Mikrococosmos¹³³, Menelas¹³⁴, Physsys¹³⁵.

E2) Ferramentas

Como exemplo de plataformas para desenvolvimento de ontologias, pode-se citar CODE4 (SKUCE, 1995), VOID (SCHREIBER; TERPSTRA; SISYPHUS, 1995), CO4 (EUZENAT, 1996), IKARUS (SKUCE, 1996), Ontolândia (FARQUHAR; FIKES; RICE, 1996), Ontosaurus (SWARTOUT et al., 1996), GKB-Editor (PALEY; KARP, 1997), JOE (MAHALINGAM; HUHNS, 1997), APECKS (TENNISON; SHADBOLT, 1998), OilEd (HORROCKS; SATTler; TOBIES, 1999), OntoEdit (MAEDCHE et al., 2000), OCM (KALFOGLOU et al, 2001), Protegé-2000 (NOY; FERGERSON; MUSEN, 2000), WebODE (ARPÍREZ et al., 2001) e WebOnto (DOMINGUE, 2001).

Exemplos de ferramentas de anotação¹³⁶ são: OntoMarkup Annotation Tool,

¹²⁸ Ver seção 4.2.4.

¹²⁹ Maiores detalhes em <<http://www.cs.utexas.edu/users/mfkb/related.html#commonkads>>.

¹³⁰ Ver item 4.3.1.

¹³¹ Maiores detalhes em <<http://www.wis.cs.utwente.nl:8080/kbs/>>.

¹³² Maiores detalhes em <<http://ksi.cpsc.ucalgary.ca/KAW/KAW98/gangemi/>>.

¹³³ Maiores detalhes em <<http://crl.nmsu.edu/Research/Projects/mikro>>.

¹³⁴ Maiores detalhes em <<http://www.biomath.jussieu.fr/~pz/Menelas/>>.

¹³⁵ Maiores detalhes em <<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/borst/node2.html>>.

¹³⁶ O princípio dessas ferramentas é proporcionar ao autor de um documento digital a possibilidade de “marcar” o documento com metadados.

OntoMat¹³⁷, OntoAnnotate¹³⁸, SHOE-Knowledge annotation¹³⁹, UBOT DAML annotation¹⁴⁰.

Exemplos de ferramentas para aprendizado sobre ontologias: ASIUM¹⁴¹, CORPORUM-Ontobuilder¹⁴², LTG Text processing wordbench¹⁴³, Text-to-Onto¹⁴⁴.

E3) Linguagens

Algumas das linguagens mais conhecidas são: *CycL* (LENAT; GUHA, 1990), *Flogic* (KIFER; LAUSEN; WU, 1990), *LOOM* (BRILL, 1993), *CARIN* (LEVY; ROUSSET, 1996), *GRAIL* (RECTOR et al., 1997), *Ontolíngua* (CHAUDHRI et al., 1998), *OCML* (DOMINGUE; MOTTA; CORCHO, 1999; CHAUDHRI; KARP; THOMERE, 1999), *OML* (KENT, 1999), *RDF* (LASSILA; SWICK, 1999), *NKRL* (BERTINO; BARBARA; ZARRI, 1999), *SHOE* (HEFLIN; HENDLER, 2000), *XOL* (KARP, 1997), *OIL* (Fensel et al., 2000), *DAML + OIL* (HORROCKS et al., 2001).

E4) Projetos que fazem uso de ontologias

Projeto	Breve descrição
<i>CoMMA</i>	Considera a memória corporativa como uma “Web semântica corporativa” e o resultado é uma ontologia; utiliza agentes inteligentes para construir uma arquitetura de informação distribuída e agentes locais que através dos quais os usuários acessam os recursos disponíveis. <i>CoMMA-Corporate Memory Management through Agents</i> (GANDON, 2001).
<i>Marchmont O.y Semantic Search Service</i>	Projeto criado pelo governo inglês para incentivar a educação continuada; trata-se de um portal onde são construídas ontologias associadas a educação. Estas ontologias indexam um banco de dados de “melhores práticas” com sumários sobre aprendizado e educação. <i>Marchmont Observatory Semantic Search Service</i> (DOMINGUE et al., 1998).
<i>MGT</i>	Ferramenta para uso na Web que constrói uma hierarquia de ontologias médicas e de suporte, as quais podem ser integradas ao banco de dados de um paciente. <i>MGT-Medical Guideline Technology</i> (HASMAN et al., 1999).
<i>MyPlanet</i>	Serviço personalizado para a Web, onde o usuário submete um <i>e-mail</i> sobre seus interesses de pesquisa; este <i>e-mail</i> é adaptado a estruturas de uma ontologia, possibilitando que seja produzida uma página da Web; os usuários são avisados sobre o assuntos de seu interesse. <i>MyPlanet</i> (KALFOGLOU, 2001).
<i>PatMan</i>	Projeto da área de saúde que provê suporte à questões médicas e administrativas no gerenciamento de pacientes; um fórum permite discussões e utiliza um motor de consultas baseado em ontologias; possibilita recuperar conhecimento sobre determinado assunto médico. <i>PatMan</i> (MOTTA, BUCKINGHAM-SHUM; DOMINGUE, 2000).
<i>PlanetOnto</i>	O <i>Kmi- Knowledge Media Institute</i> é um servidor de notícias para a Web que facilita a comunicação em empresas; a medida que o arquivo de notícias no servidor torna-se maior, ocorrem problemas de gestão (semântica, recuperação de informação, personalização, etc); Um conjunto de ferramentas chamado <i>PlanetOnto</i> , permite formalizar os documentos utilizando-se ontologias. <i>PlanetOnto</i> (DOMINGUE; MOTTA, 1999).

Figura E6 – Projetos relacionados à gestão do conhecimento

¹³⁷ Maiores informações em <<http://www.aifb.uni-karlsruhe.de/mailman/listinfo/ontomat>>.

¹³⁸ Maiores informações em <<http://www.ontoprise.de/ontocollect.htm>>.

¹³⁹ Para maiores detalhes, ver seção 5.2.10.

¹⁴⁰ Maiores detalhes em <<http://ubot.lockheedmartin.com/ubot/>>.

¹⁴¹ Maiores detalhes em <<http://www-ai.ijs.si/~ilpnet2/systems/asium.html>>.

¹⁴² Maiores informações em <http://www.cognit.no/home_multi/html/index.asp?SideID=274>.

¹⁴³ Maiores detalhes em <<http://www.ltg.ed.ac.uk/%7Emikheev/workbench.html>>.

¹⁴⁴ Maiores detalhes em <<http://ontoserver.aifb.uni-karlsruhe.de/texttoonto/>>.

Projeto	Breve descrição
<i>MKBEEEM</i>	Proporciona capacidades multi-linguísticas ao fluxo de informação em portais B2C ¹⁴⁵ ; permite manutenção semi-automática de catálogos de produtos, tradução automática e interpretação da linguagem natural de requisições dos usuários; interatividade é possível através do uso de serviços de navegação e entradas em linguagem natural. <i>MKBEEEM-Multilingual Knowledge Based European Electronic Marketplace</i> (LEGER et al., 2000).
<i>SMART-EC</i>	Plataforma de intermediação baseada em ontologias que fornece serviços para a Internet como: troca de informações entre provedores e usuários finais, definição e implementação de ciclo de vida de serviços e a possibilidade de compras em diversos sites a partir de interface única. <i>Smart-EC Support for Mediation And brokering for Electronic Commerce</i> (VÁZQUEZ; VALERA; BELLIDO, 2001).

Figura E7 – Projetos relacionados à comércio eletrônico

Projeto	Breve descrição
<i>Oncoterm</i>	Facilita a tradução de textos médicos sobre oncologia através de uma ontologia baseada em textos especializados e dicionários médicos; os conceitos são organizados em categorias e representados por esquemas. <i>Onco-term</i> (MORENO; HERNÁNDEZ, 2000).
<i>Gazelle</i>	Sistema que traduz textos japoneses, árabes e espanhóis para o inglês; inclui processamento e análise semântica das várias línguas, geração de sentenças em inglês, construção de ontologias interlíngua e criação de léxicos para japonês, árabe, espanhol e inglês. <i>GAZELLE</i> (GERMANN, 1998).
<i>Mikrokosmos</i>	Mecanismo de tradução em que se traduz textos em linguagem natural para textos no formato interlíngua (linguagem neutra, chamada <i>TMR-text meaning representation</i>); a ligação da ontologia com o TMR é feita através de um léxico, onde os significados dos itens são definidos por sua correlação com conceitos ontológicos. <i>Mikrokosmos</i> (BEALE; NIRENBURG; MAHESH, 1995).
<i>PANGLOSS</i>	Sistema de tradução que um sistema baseado em exemplos e um sistema de transferência léxica que trabalham em paralelo propondo traduções das entradas; o resultado ou tradução final é selecionada por um modelo estatístico. <i>PANGLOSS</i> (FREDERKING, 1994).
<i>KPML - Komet-Penman MultiLingual</i>	Projetado para o desenvolvimento de gramáticas e para a geração de linguagens naturais; oferece um ambiente de desenvolvimento gráfico para a construção, manutenção e uso de gramáticas; gramáticas para diversas línguas (inglês, alemão, holandês, chinês, espanhol, russo, búlgaro e checo).são desenvolvidas com o KPML <i>KPML Komet-Penman Multilingual</i> (BATEMAN, 1996).
<i>Ontogeneration</i>	Sistema para geração de textos em espanhol no domínio da química, que utiliza linguagens naturais para responder à consultas sobre grupos químicos, elementos, propriedades, etc; utiliza uma ontologia da química (<i>Chemicals</i>), uma ontologia lingüística e uma gramática em espanhol; como resultado o usuário pode consultar todas as informações da ontologia em espanhol. <i>ONTOGENERATION</i> (AGUADO et al., 1998).
<i>Penman</i>	Aplicativo para geração de sentenças em linguagem natural a partir de uma entrada não lingüística; aceita várias notações de entrada e foi projetado para o uso por pessoas com vários graus de sofisticação lingüística e computacional. <i>Komet</i> (TEICH; BATEMAN, 1993).
<i>TechDoc</i>	Protótipo que gera documentos técnicos multi-linguísticos (inglês, alemão e francês) a partir de uma representação independente construída pela análise comparativa de partes dos manuais (técnicos e lingüísticos) disponíveis em diferentes línguas. <i>TeckDoc</i> (RÖSNER; STEDE, 1994).

Figura E8 – Projetos relacionados ao processamento de linguagens naturais

¹⁴⁵

Business-to-consumer – expressão utilizada no contexto da Internet e que significa de transação de “empresa para consumidor final”.

Projeto	Breve descrição
<i>OntoSeek</i>	Sistema de agentes inteligentes projetado para recuperar informações de catálogos de produtos <i>on-line</i> ; Combina um mecanismo de <i>matching</i> ¹⁴⁶ baseado em ontologias para tratar o conteúdo e um formalismo para representação. <i>ONTOSEEK</i> (BORGO et al., 1997).
<i>WebKB-2</i>	Servidor que permite a usuários da Web recuperar e adicionar conhecimento em uma base compartilhada; permite a publicação de informações automaticamente recuperáveis e comparáveis com as de outros usuários. <i>WebKB-2</i> (MARTIN; EKLUND, 2001).
<i>C-Web - Community Web</i>	Dá suporte a comunidades da web que tem em comum conhecimento formalizado utilizado em suas atividades; conseguir um ponto de acesso único para as várias fontes de informação das comunidades é a limitação atual da ferramenta. <i>C-Web</i> (ALEXAKI et al., 2002).
<i>SEAL - Semantic Portal</i>	Abordagem genérica baseada em ontologias para o desenvolvimento de portais semânticos; explora o aspecto semântico através do fornecimento e acesso à informações em um portal. <i>SEAL</i> (MAEDCHE et al., 2001).
<i>Sum Microsystems Conceptual Indexing</i>	Ferramenta baseada em uma taxonomia conceitual, com base de conhecimento do tipo KL-One com cerca de 3000 regras; sistema de indexação conceitual automático; relacionamentos “tipo-de”; suporta textos e páginas da Web.
<i>Verity Search</i>	Ferramenta comercial de busca em texto completo para portais de <i>e-business</i> ; atua em Internet, <i>intranets</i> , <i>extranets</i> e bancos de dados; a estrutura conceitual do tipo esquema de classificação é representada por linguagem proprietária, que atua em um grafo semântico com pesos; os relacionamentos ocorrem entre palavras com a atribuição de pesos; as regras são construídas manualmente por especialistas.
<i>Mindwave</i>	Ferramenta comercial para recuperação baseada em conteúdo, que atua na Internet, <i>intranets</i> e <i>extranets</i> ; trabalha com modelos de previsão ¹⁴⁷ e redes neurais além de modelo de vetores (tecnologia proprietária); os relacionamentos na estrutura conceitual são obtidos por categorização e aprendizado automático e fazem associações de similaridade entre vetores de contexto.
<i>LexiGuide</i>	Ferramenta comercial para busca em Internet e <i>intranets</i> baseada em agentes inteligentes; a estrutura conceitual é baseada em uma rede semântica lingüística; os relacionamentos são ligações entre palavras e conceitos semânticos; a estrutura é criada por processamento de linguagem natural.
<i>Excalibur RetrievalWare</i>	Ferramenta comercial para busca na Internet, baseada em agentes inteligentes; rede semântica de conceitos com uma base de conhecimento de 5000 significados de palavras e cerca de 1,6 milhões de relacionamentos; os relacionamentos são proporcionados por ligações entre conceitos para dicionários e <i>thesaurus</i> ; a estrutura é criada por processamento de linguagens naturais;
<i>ReadWare ConSearch</i>	Ferramenta comercial para buscas na Internet e <i>intranets</i> , baseada em agentes inteligentes, a estrutura é construída a partir dos radicais das palavras e é representada por rede semântica; os relacionamentos são baseados em matrizes conceitos x relacionamentos; a estrutura é criada por análise automática de textos.
<i>Tunderstone MetaMorph</i>	Ferramenta comercial para busca interna em <i>sites</i> ; a estrutura conceitual é um <i>thesaurus</i> , representado por uma rede lingüística; não utiliza técnicas de indexação; as relações são baseadas em associação de palavras e operações lógicas em conceitos; a estrutura é criada por processamento de linguagem natural.
<i>Tunderstone Webnator</i>	Ferramenta comercial para busca de documentos em <i>intranets</i> ; a estrutura conceitual é baseada em <i>thesaurus</i> e representada por um rede lingüística; as relações baseiam-se nas associações entre palavras e frases; possibilita indexação comum entre vários <i>sites</i> .
<i>DioWebSearch</i>	Ferramenta comercial para busca contextual, que atua na Internet e <i>intranets</i> ; a estrutura conceitual é uma rede semântica lingüística representada por uma rede lingüística contextual-morfológica; as relações na estrutura se baseiam na associação de conceitos e palavras; a estrutura é criada por processamento de linguagens naturais.
<i>FOML (formal ontology markup Language)</i>	Trata-se de uma linguagem de marcação para a construção de ontologias (OGATA, 2001).

Figura E9 – Projetos relacionados à recuperação da informação na Web

¹⁴⁶ Algoritmos de casamento de padrão.

¹⁴⁷ A previsão equivale a uma classificação, em que os registros são classificados de acordo com algum comportamento futuro previsto ou valores estimados.

Projeto	Breve descrição
<i>RichODL</i>	Ambiente de aprendizado na Web desenvolvido pela <i>Open University</i> ¹⁴⁸ para treinar de estudantes na modelagem e simulação de ambientes dinâmicos; ontologias são usadas para descrever o domínio físico dos sistemas modelados além de suas correlações e comportamento. <i>RichODL</i> (ZDRAHAL et al., 2000).
<i>Smartrainer</i>	Sistema de treinamento automático que inclui simuladores para sistemas elétricos; o objetivo é treinar funcionários de empresas de energia na recuperação de acidentes em subestações elétricas. <i>SmartTrainer</i> (JIN et al., 1997).
<i>SchoolOnto</i> <i>Scholarly Ontologies Project</i>	Biblioteca digital baseada em ontologias que possibilita a interpretação dos domínios.; auxilia na modelagem de campos de pesquisas, os quais são dinâmicos e carecem de ferramentas para uniformização de inconsistências; permite ao usuário descrever e discutir, através de uma rede semântica, a contribuição do documento para a literatura da área. <i>ScholOnto</i> (SHUM; MOTTA; DOMINGUE, 2000).

Figura E10 – Projetos relacionados educação

Como exemplos de aplicações que, de alguma forma, utilizam ontologias, apresentam-se ainda: *Generating InStructional Text (GIST)* (STOCK, 1996), *Ariadne* (ARENS; KNOBLOCK; SHEN, 1996), *OBSERVER* (MENA et al., 1996), *PICSEL* (GOASDOUÉ; LATTES; ROUSSET, 1998), *ITEM-Search Engine* (GONZALO; PEÑAS; VERDEJO, 1999), *ALT-JE* (OGURA; BOND; OOHAMA, 1999).

¹⁴⁸

<<http://www.open.ac.uk/>>.

Anexo F – Lista de classes, propriedades e comentários de algumas ontologias sobre registros bibliográficos

- (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>> (21 campos mais utilizados do padrão MARC)¹⁴⁹
 (6) <<http://orlando.drc.com/SemanticWeb/Topics/Ontology/Ontologies.htm>>(ontologia padrão *Dublin Core*).

URL	Campo da ontologia	Função na ontologia	Comentários feitos no corpo da ontologia
(5)	<i>Added entry-- Personal name</i>	Campo MARC	Indica o tipo de nome pessoal (sobrenome, da família, primeiro nome) ou se nenhuma informação é fornecida (co-autores, editores, etc)
(5)	<i>Added entry - Corporate name</i>	Campo MARC	Indica o tipo de nome da organização (nome invertido, pessoa jurídica) ou se nenhuma informação é fornecida.
(5)	<i>Added entry - Uncontrolled related/analytical title</i>	Campo MARC	Indica número de caracteres não preenchidos em um campo (por exemplo, espaços).
(3)	<i>Article</i>	Classe	Subclasse de <i>Publication-</i>
(4)	<i>Article</i>	Classe	Um artigo de um jornal ou revista
(2)	<i>Author</i>	Propriedade	cardinalidade mínima 1
(1)	<i>Author</i>	Propriedade	Uma pessoa ou grupo responsável para conteúdo intelectual ou artístico de um documento.
(3)	<i>Author</i>	Propriedade	Domain= publication; range=literal
(6)	<i>Author/Creator</i>	Propriedade	Uma entidade responsável por criar o conteúdo para o recurso, podendo ser pessoa, organização ou serviço.
(1)	<i>Available</i>	Propriedade	Fonte comercial na qual o documento citado pode ser obtido ou a localização onde o documento pode ser encontrado.
(5)	<i>Bibliography, etc. note</i>	Campo MARC	Tipo de nota de conteúdo (completa ou incompleta, de volumes ainda não publicados ou conteúdo parcial)
(1)	<i>Bibliographic</i>	Classe	Inclui elementos que podem ser utilizados com categorias de dados em citações bibliográficas;
(1)	<i>BibliographicEntry</i>	Classe	Uma entrada em que o conteúdo consiste de uma citação bibliográfica padrão.
(1)	<i>BibliographicEntryFor</i>	Propriedade	-
(4)	<i>Bibtex_Entry</i>	Classe	Um entrada para um banco de dados bibtex
(1)	<i>Book</i>	Propriedade	Publicação completa em um volume ou que será concluída com um número finito de volumes.
(2)	<i>Book</i>	Classe	Um livro com um editor específico
(3)	<i>Book</i>	Classe	Subclasse de <i>publication</i>
(4)	<i>Book</i>	Classe	Um livro com um editor definido
(4)	<i>Booklet</i>	Classe	Um trabalho que é impresso mas sem o nome de um editor
(2)	<i>Booktitle</i>	Propriedade	Cardinalidade 1
(3)	<i>Booktitle</i>	Propriedade	Domain= Publication; range= Literal"
(5)	<i>Cataloging source</i>	Campo MARC	Agência de catalogação original, agência de transcrição
(1)	<i>Categoria</i>	Propriedade	tipo da publicação
(2)	<i>Colection</i>	Propriedade	uma coleção é um livro com artigos de autores diferentes.
(4)	<i>Conference</i>	Classe	-
(1)	<i>City</i>	Propriedade	cidade e estado em que o documento é publicado;
(1)	<i>Citation</i>	Propriedade	referência a um livro, artigo ou outra fonte
(1)	<i>Coden</i>	Propriedade	Um código alfa-numérico único atribuído a uma série e usado com um identificador sem ambiguidades
(6)	<i>Contributor</i>	Propriedade	Entidade responsável por fazer contribuições ao conteúdo do recurso, podendo ser pessoa, organização ou serviço.
(1)	<i>CodeNumber</i>	Propriedade	-
(1)	<i>CorporateAuthor:</i>	Propriedade	A organização responsável pela criação de todo ou algum conteúdo intelectual ou artístico do trabalho.
(1)	<i>Copyright:</i>	Propriedade	Direito legal e exclusivo, garantido por um período para um autor, projetistas, etc; ou outra pessoa indicada, para imprimir, publicar, executar filmar ou registrar originais de literatura, artigos e músicas.
(1)	<i>CopyrightHolder</i>	Propriedade	O nome do proprietário dos direitos autorais na época da publicação
(1)	<i>Country</i>	Propriedade	País em que o documento é publicado;
(6)	<i>coverage</i>	propriedade	A extensão ou o escopo do conteúdo do recurso.

149

Alguns campos, que no processo de seleção, foi identificado como idêntico em outras ontologias, pode ter sido omitido.

(6)	<i>Date</i>	propriedade	Vai ser associada co a criação ou disponibilidade do recurso.
(1)	<i>DateOfPublication</i>	Propriedade	Indicação do ano e se necessário, do mês e dia da publicação.
(6)	<i>Description</i>	propriedade	Uma descrição do conteúdo do recurso.
(1)	<i>Editor</i>	Propriedade	Organização ou pessoa responsável pela preparação da publicação de um documento, a partir do ponto de vista de seu conteúdo intelectual
(3)	<i>Editor</i>	Propriedade	<i>Domain= publication; range= Literal</i>
(1)	<i>Extent</i>	Propriedade	Número de páginas de um trabalho
(1)	<i>Edition</i>	Propriedade	O conjunto de cópias de um documento de uma composição ou de uma única cópia utilizada como original;
(5)	<i>Edition statement</i>	Campo MARC	-
(1)	<i>FirstName</i>	Propriedade	O nome de um indivíduo
(6)	<i>Format</i>	propriedade	Podem incluir o tipo de mídia ou as dimensões do recurso; manifestação física ou digital do recurso.
(5)	<i>Formatted contents note</i>	Campo MARC	Tipo de nota de conteúdo (conteúdo completo, incompleto ou parcial)
(5)	<i>General note</i>	Campo MARC	Usado quando um campo de notas não especializado foi definido para a informação (por exemplo, notas que dizem respeito ao índice, a fonte do título, o escopo do item, etc)
(1)	<i>Glossary</i>	Propriedade	Lista terminológica de um pequeno dicionário contendo a terminologia de um campo específico ou de campos relacionados.
(4)	<i>hasPublication_Type</i>	Propriedade	-
(4)	<i>hasCiteKey</i>	Propriedade	-
(4)	<i>hasAddress</i>	Propriedade	-
(4)	<i>hasAnnote</i>	Propriedade	-
(4)	<i>hasAuthor</i>	Propriedade	-
(4)	<i>hasBooktitle</i>	Propriedade	-
(4)	<i>hasChapter</i>	Propriedade	-
(4)	<i>hasCrossref</i>	Propriedade	-
(4)	<i>hasEdition</i>	Propriedade	-
(4)	<i>hasEditor</i>	Propriedade	-
(4)	<i>hasHowpublished</i>	Propriedade	-
(4)	<i>hasInstitution</i>	Propriedade	-
(4)	<i>hasJournal</i>	Propriedade	-
(4)	<i>hasKey</i>	Propriedade	-
(4)	<i>hasMonth</i>	Propriedade	-
(4)	<i>hasNote</i>	Propriedade	-
(4)	<i>hasNumber</i>	Propriedade	-
(4)	<i>hasOrganization</i>	Propriedade	-
(4)	<i>hasPages</i>	Propriedade	-
(4)	<i>hasPublisher</i>	Propriedade	-
(4)	<i>hasSchool</i>	Propriedade	-
(4)	<i>hasSeries</i>	Propriedade	-
(4)	<i>hasTitle</i>	Propriedade	-
(4)	<i>hasType</i>	Propriedade	-
(4)	<i>hasVolume</i>	Propriedade	-
(4)	<i>hasYear</i>	Propriedade	-
(4)	<i>InBook</i>	Classe	Parte de um livro, em geral, sem título
(3)	<i>Incollection</i>	Classe	Subclasse de <i>publication</i>
(4)	<i>InCollection</i>	Propriedade	Uma parte de um livro com seu título.
(2)	<i>InProceedings</i>	Propriedade	um artigo que tem como característica esta em um <i>proceedings</i> .
(3)	<i>Inproceedings</i>	Classe	Subclasse de <i>publication</i>
(4)	<i>InProceedings</i>	Classe	Um artigo em uma conferencia.
(2)	<i>Institution</i>	Propriedade	Cardinalidade 1
(3)	<i>Institution</i>	Propriedade	<i>domain Publication; range Literal</i>
(1)	<i>Isbn</i>	Propriedade	Número único e padrão, internacionalmente conhecido, atribuído a cada edição de um livro ou outras publicações, para identificação;
(5)	<i>International Standard Book Number</i>	Campo MARC	Número padrão para identificação de uma publicação.
(1)	<i>Issn</i>	Propriedade	Número único e padrão, internacionalmente conhecido, atribuído a cada edição de um livro ou outras publicações, para identificação
(1)	<i>IssueNumber</i>	Propriedade	Número de um documento publicado durante um certo período de tempo
(2)	<i>Jornal</i>	Propriedade	Cardinalidade 1
(3)	<i>Journal</i>	Propriedade	<i>domain= Publication; range= Literal</i>
(5)	<i>Language</i>	Campo MARC	Linguagem do conteúdo do recurso.
(4)	<i>Literal</i>	Classe	-
(5)	<i>Main entry -- Personal name</i>	Campo MARC	Tipo de nome (sobrenome, nome próprio, nome de família) do autor principal;
(5)	<i>Main entry -- Uniform title</i>	Campo MARC	Pode indicar um título, parte de um trabalho, linguagem do trabalho, versão, data do trabalho.
(2)	<i>MasterThesis</i>	classe	-
(4)	<i>MastersThesis</i>	Propriedade	-
(4)	<i>Manual</i>	Classe	Documentação técnica
(4)	<i>Misc</i>	Classe	Um conjunto de tipos de publicações usado quando nada se combina
(3)	<i>Note</i>	propriedade	<i>domain= Publication; range= Literal</i>
(3)	<i>Number</i>	propriedade	<i>domain= Publication; range= Literal</i>
(1)	<i>OrganizationName</i>	Propriedade	Nome de uma associação, negócio ou outro entidade estruturada.

(1)	<i>OtherInformation</i>	Propriedade	Informações genéricas usadas para itens não tratados na lista de elementos; usado na terminologia de gerenciamento de banco de dados para elemento terminológicos em geral, não tratados em bibliotecas.
(1)	<i>Page</i>	Propriedade	Um dos lados de uma folha de um documento
(3)	<i>Pages</i>	Propriedades	<i>domain= Publication; range= Literal</i>
(5)	<i>Physical description</i>	Campo MARC	Indica número de páginas, dimensões, material complementar (por exemplo, guia do professor)
(2)	<i>PhdThesis</i>	Classe	-
(4)	<i>PhdThesis</i>	Classe	-
(2)	<i>Proceeding</i>	Classe	Uma coleção de artigos de uma conferência
(4)	<i>Proceedings</i>	Classe	<i>Proceedings</i> de uma conferência
(5)	<i>Publication, distribution</i>	Campo MARC	Lugar, nome e data da publicação e distribuição
(2)	<i>Publication</i>	Classe	uma classe de publicações de alto nível
(3)	<i>Publication</i>	classe	Todos os termos aqui vem do BiBtex
(1)	<i>PublicationCategory</i>	Propriedade	-
(1)	<i>PublicationDate</i>	Propriedade	-
(6)	<i>Publisher</i>	Propriedade	Uma entidade responsável por tomar o recurso disponível, podendo ser uma pessoa, empresa ou serviço.
(2)	<i>Publisher</i>	Propriedade	cardinalidade 1
(3)	<i>Publisher</i>	Propriedade	<i>domain= Publication; range= Literal</i>
(1)	<i>PublisherName</i>	Propriedade	Nome da pessoa ou organização responsável pela produção e disseminação de um documento.
(1)	<i>ReferenceFor</i>	Propriedade	-
(6)	<i>Relation</i>	Propriedade	Uma referência a um recurso relacionado.
(1)	<i>ReportIdentifier</i>	Propriedade	Designação alfa-numérica única que identifica um relatório
(6)	<i>Resource</i>	Propriedade	Um recurso unívoco para um recurso dentro de um dado contexto.
(1)	<i>Role</i>	Propriedade	Principal função de alguém que originou um documento
(6)	<i>Rights</i>	Propriedade	Informação sobre direitos autorais de um recurso.
(2)	<i>Scholl</i>	Propriedade	Cardinalidade 1
(5)	<i>Series added entry - Uniform title</i>	Campo MARC	Pode indicar título ou número do volume.
(5)	<i>Series added entry - Personal name</i>	Campo MARC	Indica tipo de nome pessoal (sobrenome, nome da família, primeiro nome) ou datas associadas com o nome, volumes, forma completa do nome, etc.
(5)	<i>Series statement / Added entry--Title</i>	Campo MARC	Pode representar título ou número do volume.
(1)	<i>SerialTitle</i>	Propriedade	O título de uma série
(6)	<i>Source</i>	Propriedade	A referência para um outro recurso do qual o presente recurso é derivado
(1)	<i>SourceFor</i>	Propriedade	-
(6)	<i>Subject</i>	Propriedade	O tópico do conteúdo do recurso, em geral, expresso por palavras chave.
(5)	<i>Subject added entry - Geographic name</i>	Campo MARC	Para sistema de cabeçalho de assuntos ou <i>thesaurus</i> ; indica nome geográfico e sub-divisão geral, cronológica ou geográfica.
(5)	<i>Subject added entry -- Topical term</i>	Campo MARC	Para sistema de cabeçalho de assuntos ou <i>thesaurus</i> (indica a lista em que o documento é usado; e sub-divisão geral, cronológica ou geográfica.
(5)	<i>Subject added entry -- Corporate name</i>	Campo MARC	Indica tipo de nome da organização (nome invertido, nome pessoa jurídica); para sistema de cabeçalho de assuntos ou <i>thesaurus</i>
(5)	<i>Subject added entry -- Personal name</i>	Campo MARC	Tipo de nome pessoal (sobrenome, nome próprio, nome da família); para sistema de cabeçalho de assuntos ou <i>thesaurus</i> .
(5)	<i>Summary, etc. note</i>	Campo MARC	Indica sumário, resumo ou anotação.
(1)	<i>Surname</i>	Propriedade	Nome hereditário de um indivíduo;
(2)	<i>TechReport</i>	Classe	um relatório publicado por uma escola ou outra instituição, em geral numerado dentro de uma série.
(3)	<i>Techreport</i>	Propriedade	Subclasse de <i>publication</i>
(4)	<i>TechReport</i>	Classe	Um relatório publicado por uma escola ou outra instituição
(6)	<i>Title</i>	Propriedade	O nome dado para o recurso
(2)	<i>Thesis</i>	Classe	-
(1)	<i>Title</i>	Propriedade	Palavra ou frase única, que em geral aparece no documento, pelo qual se refere a ele, que podem ser usado para identificá-lo e distribuí-lo de outros documentos;
(2)	<i>Title</i>	Propriedade	cardinalidade 1
(3)	<i>Title</i>	Propriedade	-
(5)	<i>Title Statement</i>	Campo MARC	Indica título próprio, meio (quando não é livro), nome de parte ou seção, subtítulos, etc.
(5)	<i>Uniform title</i>	Campo MARC	Pode indicar título que é apresentado ou impresso, língua e data do trabalho.
(4)	<i>Unpublished</i>	Classe	Um artigo com o autor e título, mas não publicado formalmente.
(5)	<i>Varying form of title</i>	Campo MARC	Pode indicar parte do título, título paralelo, título de capa.
(3)	<i>Volume</i>	Propriedade	-
(1)	<i>VolumeIdentifier:</i>	Propriedade	Identificador para um unidade de material que reúne um certo número de unidades que forma um conjunto parcial ou total.
(3)	<i>Year</i>	Propriedade	-
(2)	<i>Year</i>	Propriedade	Cardinalidade 1

Figura F1 – Lista de campos relacionados por ordem alfabética

Termos da lista utilizados	Classe definida	Descrição	Subclasses
Coverage (6)	<u>Domain</u>	Domínio do conhecimento que se está estudando. Diz respeito a acervo bibliográfico.	<u>Collections</u> : todas as instâncias da subclasse <i>collections</i> são também instâncias de <i>domain</i> . Por exemplo, instâncias de <i>collection</i> são “base USP”, “base UFMG”, etc, as quais também são instâncias de <i>domain</i> . Várias <i>collections</i> fazem parte de um <i>domain</i> .
-	<u>Collections</u>	Coleção de publicações de um domínio do conhecimento. Diz respeito a acervo bibliográfico.	<u>Publications</u> : todas as instâncias da subclasse <i>publications</i> são também instâncias de <i>collections</i> . Por exemplo, instâncias de <i>publications</i> são <i>articles</i> de nome X, <i>books</i> de nome Y, etc, os quais também são instâncias de <i>collections</i> . Vários <i>publications</i> , por exemplo livros do assunto Z fazem parte de uma <i>collection</i> .
Publication (5) Publication, Distribution (5) Publication (2)	<u>Publications</u>	Documentos relativos ao domínio do conhecimento que estão disponíveis à comunidade científica e ao público em geral, através de sua publicação. Diz respeito a acervo bibliográfico.	<u>Articles, Books, Journals, Manual, Master thesis, Miscelania, Periodics, Phd thesis, Proceedings, Tech report</u> . Por exemplo, artigo X é uma instância de <i>articles</i> e também uma instância de <i>publications</i> . Ou seja, o <i>article</i> é um tipo de <i>publication</i> . As outras subclasses são definidas de forma análoga.
Article (3) Article (4)	<u>Articles</u>	Um documento que é publicado em jornal ou revista científica. Diz respeito a acervo bibliográfico.	NT
Book (1) Book (2) Book (3) Book (4) Booklet (4)	<u>Books</u>	Uma obra publicada com editor definido. Diz respeito a acervo bibliográfico.	NT
Jornal (2) Journal (3)	<u>Journal</u>	Publicação científica de caráter informativo. Diz respeito a acervo bibliográfico.	NT
Manual (4)	<u>Manual</u>	Documento que objetiva ensinar sobre o funcionamento de um determinado processo. Diz respeito a acervo bibliográfico.	NT
Miscelania	<u>Miscelania</u>	Conjunto de tipos de <i>publications</i> em que nada se combina. Diz respeito a acervo bibliográfico.	NT
MasterThesis (2) MasterThesis (4)	<u>Master thesis</u>	Publicação que consiste de uma dissertação de mestrado. Diz respeito a acervo bibliográfico.	NT
PhdThesis (2) PhdThesis (4)	<u>Phd thesis</u>	Publicação que consiste de uma tese de doutorado. Diz respeito a acervo bibliográfico.	NT
Proceeding (2) Proceedings (4)	<u>Proceedings</u>	Um grupo de artigos de uma conferência. Diz respeito a acervo bibliográfico.	NT
TechReport (2) Techreport (3) TechReport (4)	<u>Tech report</u>	Relatório publicado por uma escola ou instituição. Diz respeito a acervo bibliográfico.	NT
-	<u>Periodic</u>	Revista especializada da área. Diz respeito a acervo bibliográfico.	NT

Figura F2 – Agrupamento semântico dos termos e definição de classes / subclasses

Classe	Slot	Facet (seção 5.5.5)
Address	Is-a	thing
Article	Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Is-published-by Has-one Has-one Has-one Has-one Is-a	Creator Title Category Date_publication Description Extent Edition Format Language Other Information Publisher Summary Citation Subject Source Journal or book Citation Copyright Extent Type Publication
Book	Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Is-published-by Has-one Has-one Has-one Has-one Has-one Is-a	Creator Title Category Date_publication Code-number Description Extent Edition Format Language Other Information Summary Citation Subject Source Publisher Journal or book Citation (cardinal. min.1) Copyright (cardinal. min.1) Extent Type Volume-number Publication
Category	Is-a	Article or book or manual or phd_thesis or master_thesis...
Citation	Has-one Has-one	Creator Publication
Code_number	Is-a	One-of ISBN
Collection	Has-one	Publication (cardinal. min.2)
Copyright	Has-one	Creator
Country	Is-a	Thing

<i>Creator</i>	<i>Is-a</i> <i>Publish-one</i>	<i>One-of enterprise or One-of person</i> <i>Article or book or manual</i>
<i>Date_publication</i>	<i>Is-a</i>	<i>Thing</i>
<i>Description</i>	<i>Is-a</i>	<i>Thing</i>
<i>Dimensions</i>	<i>Is-a</i>	<i>Thing</i>
<i>Domain</i>	<i>Has-one</i> <i>Is-a</i> <i>Has-one</i>	<i>Key</i> <i>On-of Information Science</i> <i>Collection (cardinal.min. 1)</i>
<i>Edition</i>	<i>Is-a</i>	<i>Thing</i>
<i>Extent</i>	<i>Has-one</i>	<i>Number_pages</i>
<i>Format</i>	<i>Has-one</i>	<i>One-of digital or One-of paper</i>
<i>Issue_number</i>	<i>Is-a</i>	<i>Thing</i>
<i>Journal</i>	<i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Is-published-by</i> <i>Has-one</i>	<i>Creator</i> <i>Title</i> <i>Category</i> <i>Date_publication</i> <i>Description</i> <i>Extent</i> <i>Edition</i> <i>Format</i> <i>Issue_number</i> <i>Other_information</i> <i>Copyrigh</i> <i>Citation</i> <i>Subject</i> <i>Type</i> <i>Publisher</i> <i>Summary</i> <i>Volume_number</i> <i>Language</i> <i>Source</i> <i>Publisher</i> <i>Issue-number</i>
<i>Key</i>	<i>Is-a</i>	<i>Thing</i>
<i>Language</i>	<i>Language</i>	<i>One-of portuguese</i>
<i>Manual</i>	<i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Is-published-by</i> <i>Has-one</i>	<i>Creator</i> <i>Title</i> <i>Category</i> <i>Date_publication</i> <i>Description</i> <i>Extent</i> <i>Edition</i> <i>Format</i> <i>Language</i> <i>Copyright</i> <i>Citation</i> <i>Subject</i> <i>Type</i> <i>Publisher</i> <i>Summary</i> <i>Volume_number</i> <i>Source</i> <i>Publisher</i>

<i>Master_thesis</i>	<i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i>	<i>Creator</i> <i>Title</i> <i>Date_publication</i> <i>Description</i> <i>Extent</i> <i>Format</i> <i>Other_information</i> <i>Copyrigh</i> <i>Citation</i> <i>Subject</i> <i>Type</i> <i>Summary</i> <i>Volume_number</i> <i>Language</i> <i>Source</i>
<i>Miscelania</i>	<i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Is-published-by</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Is-a</i>	<i>Creator</i> <i>Title</i> <i>Category</i> <i>Date_publication</i> <i>Description</i> <i>Extent</i> <i>Edition</i> <i>Format</i> <i>Language</i> <i>Other Information</i> <i>Publisher</i> <i>Summary</i> <i>Citation</i> <i>Subject</i> <i>Source</i> <i>Journal or book</i> <i>Citation</i> <i>Copyright</i> <i>Extent</i> <i>Type</i> <i>Publication</i>
<i>Number_pages</i>	<i>Is-a</i>	<i>Thing</i>
<i>Other_information</i>	<i>Is-a</i>	<i>Thing</i>
<i>Periodics</i>	<i>Has-one</i> <i>Is-published-by</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i>	<i>Article (cardinal.min.3)</i> <i>Publisher</i> <i>Date_publication</i> <i>Language</i> <i>Issue_number</i>
<i>Phd_Thesis</i>	<i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i> <i>Has-one</i>	<i>Creator</i> <i>Title</i> <i>Date_publication</i> <i>Description</i> <i>Extent</i> <i>Format</i> <i>Other_information</i> <i>Copyrigh</i> <i>Citation</i> <i>Subject</i>

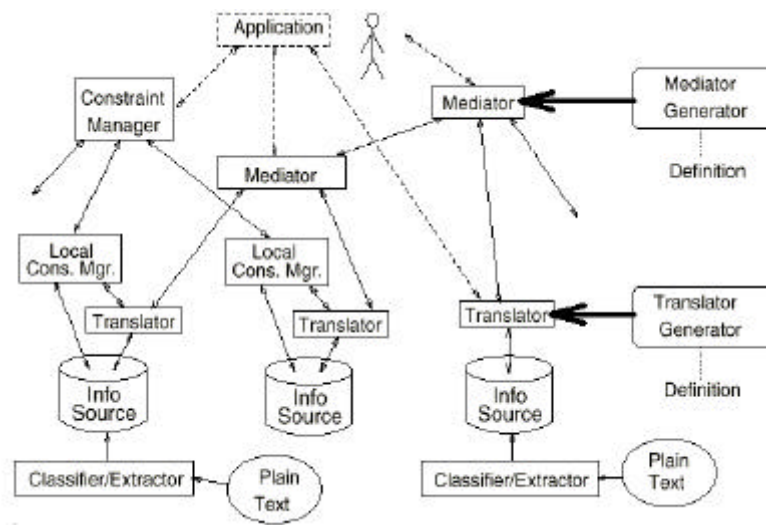
	Has-one Has-one Has-one Has-one Has-one	Type Summary Volume_number Language Source
Place_available	Is-a	One-of Internet or Journal or Book or periodcs
Proceedings	Has-one	Publication
Publication	Has-one Has-one Has-one Is-a	Place_available Category Citation (cardinal.min.1) Article or book or journal or manual or ...
Publisher	Is-a Has-one Publish-one	One-of enterprise Addrees Article or book or journal
Source	Is-a Is-a	Country (cardinal.min 1) University (cardinal.min 1)
Subject	Is-a	Thing
Summary	Is-a	Thing
Tech_report	Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Has-one Is-published-by Has-one	Creator Title Category Date_publication Description Extent Edition Format Language Copyright Citation Subject Type Publisher Summary Volume_number Source Publisher
Title	Is-a	Thing
Type	Is-a	Thing
University	Has-one Has-one	Address Country
Volume_number	Is-a	Thing

Figura F3 – Classes, seus slots e facets

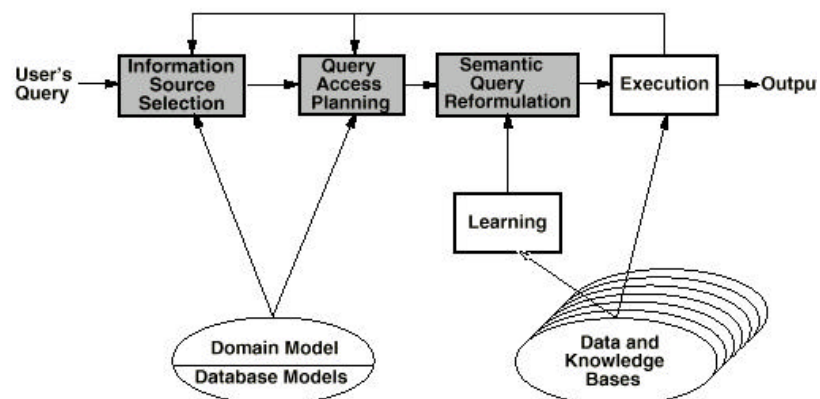
Anexo G – Arquiteturas de diversos sistemas para integração

Os sistemas apresentados no Capítulo 4, especificamente, na Seção 4.3, procuram executar integração e são baseados em ontologias. Para maior detalhamento, caso necessário, este anexo traz as figuras das arquiteturas de cada sistema apresentado:

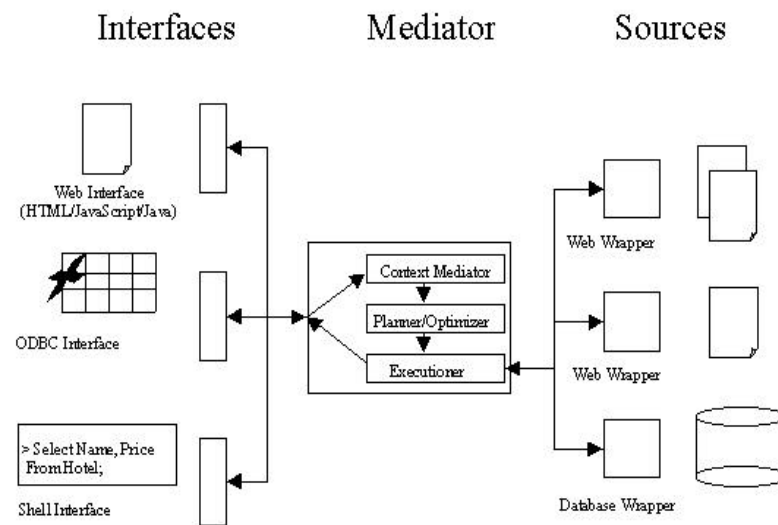
TSIMMIS



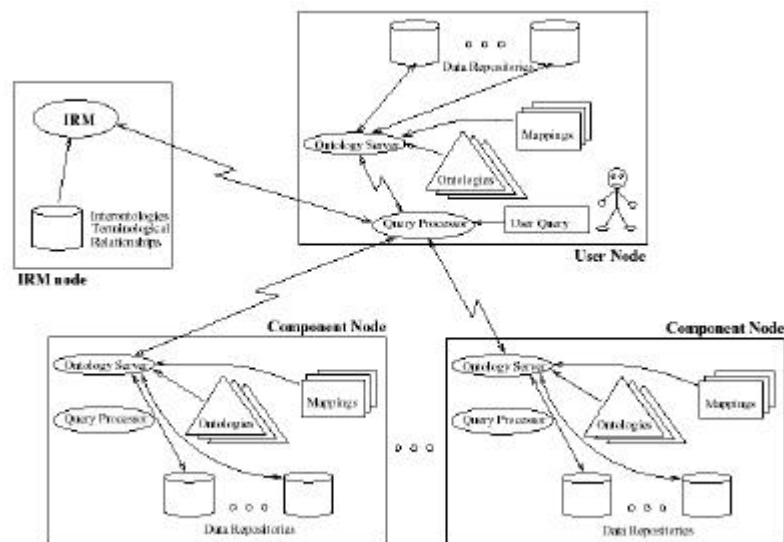
SIMS



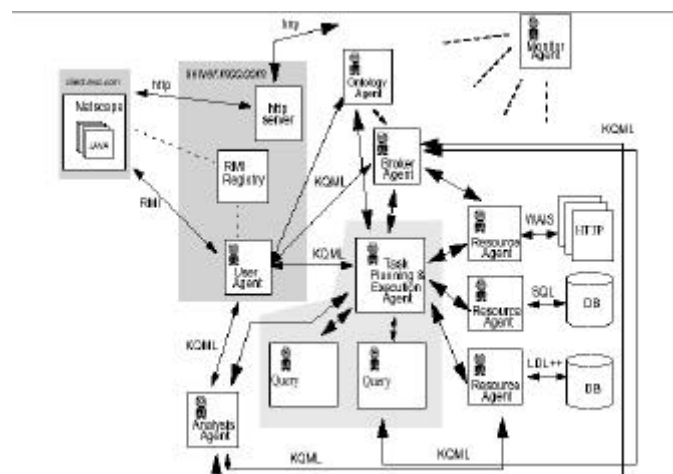
COIN



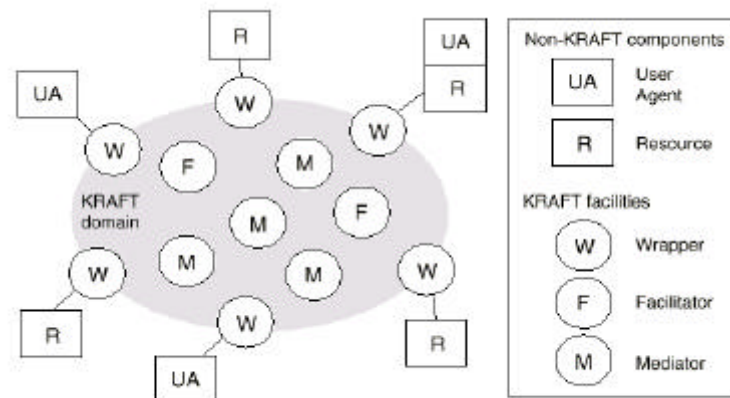
OBSERVER



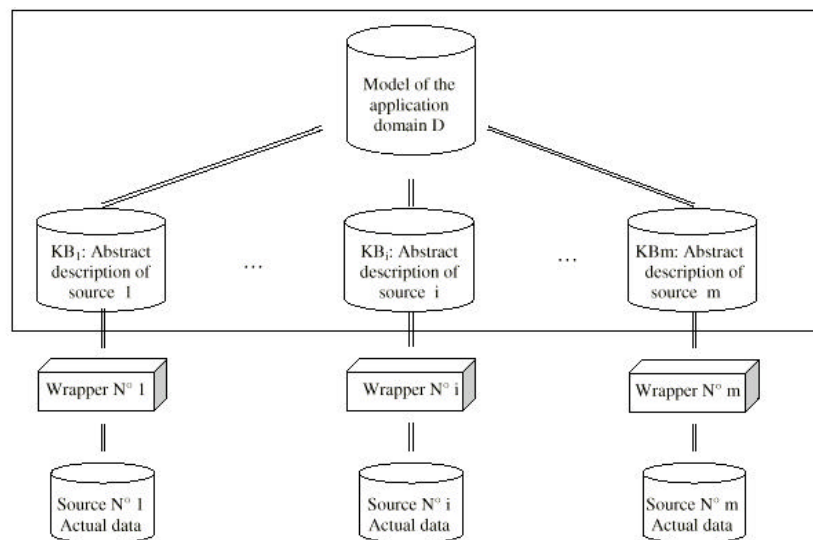
INFOSLEUTH



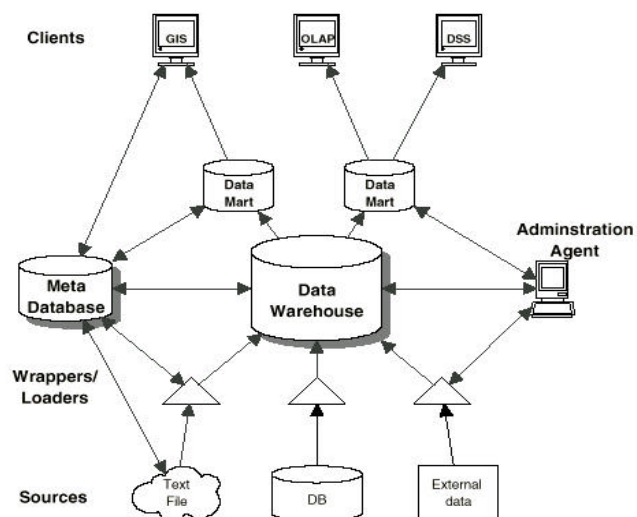
KRAFT



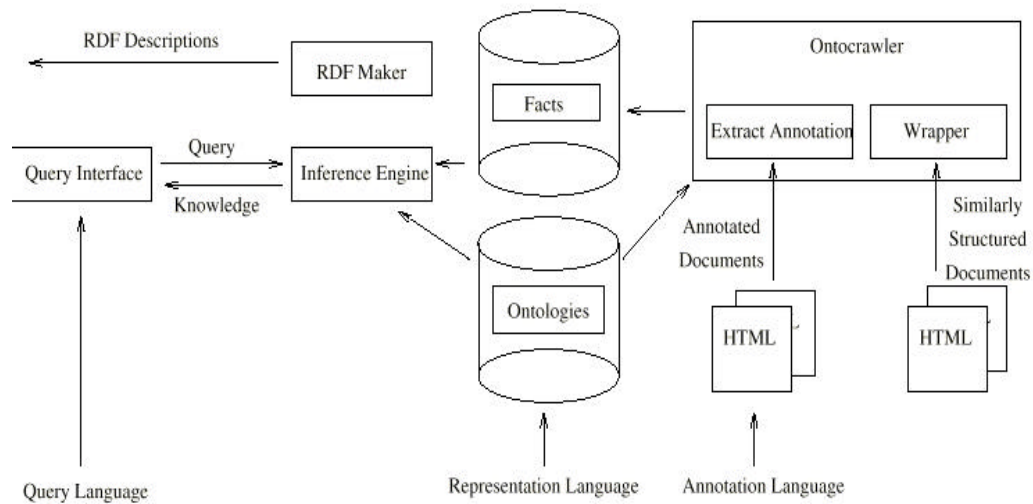
PICSEL



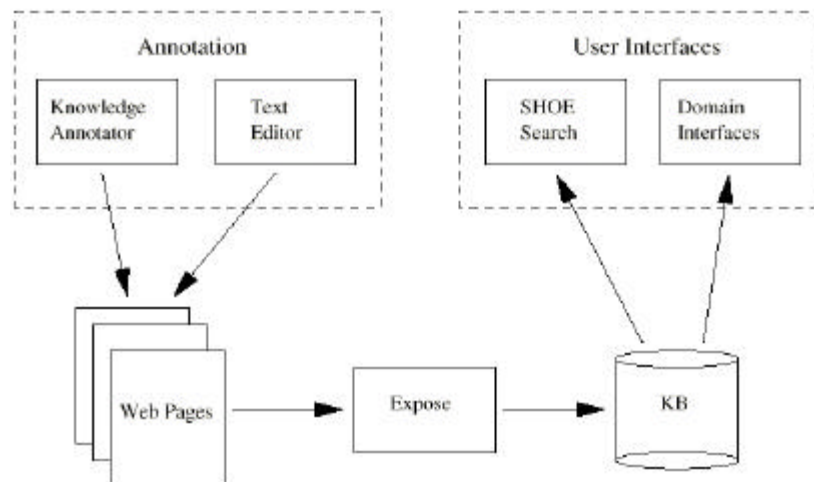
DWQ



ONTOBROKER



SHOE



BUSTER

