

Padrões de Mapeamentos Relaxados entre Ontologias Formais utilizando Contexto Local e Não-Local

Rodrigo Camelo de Oliveira
Orientadora: Virgínia V. B. Biris Brilhante

Proposta de dissertação apresentada ao Programa
de Pós-Graduação em Informática do Instituto de
Ciências Exatas da Universidade Federal do Ama-
zonas como requisito parcial para obtenção do título
de Mestre em Informática.

Manaus
Fevereiro de 2005

Padrões de Mapeamentos Relaxados entre Ontologias Formais utilizando Contexto Local e Não-Local

Rodrigo Camelo de Oliveira
rcamel@ufam.edu.br

Programa de Pós-Graduação em Informática
Universidade Federal do Amazonas

Resumo

Ontologias vêm sendo desenvolvidas e utilizadas por aplicações em diversos domínios de conhecimento de variadas formas [42, 10], como recurso de suporte à consistência e compatibilidade de informações compartilhadas. O desenvolvimento destas ontologias por diferentes equipes, normalmente norteados por diferentes objetivos e filosofias, gerou um grande número de ontologias cobrindo domínios sobrepostos e/ou relacionados. Recentemente, seguindo as diretrizes da Web Semântica, ontologias vêm progressivamente sendo utilizadas na World Wide Web onde provêem semântica para anotações em páginas web. Dada a natureza distribuída do desenvolvimento das ontologias e da World Wide Web, é importante que as anotações possuam semânticas compatíveis, o que requer integração entre as ontologias. A definição de mapeamentos entre conceitos (de diferentes ontologias) com alguma relação semântica é, uma das formas de integração entre as ontologias envolvidas. Contudo, processos atuais de integração [27] exigem intensa intervenção manual, tornando o processo tedioso, demorado e pouco reusável. Pretende-se estabelecer uma técnica que, observando o contexto das estruturas que compõem as ontologias envolvidas, identifique relações semânticas entre conceitos, classifique-as, e estabeleça mapeamentos (relaxados, caso necessário) entre estes conceitos.

1 Introdução

As tecnologias da informação, em particular a World Wide Web, aceleram a geração e a disponibilização de informação, que cresce não apenas em volume mas em diversidade de conteúdos e estruturas. Ao mesmo tempo, valorizam-se estudos multi e interdisciplinares na análise e solução de problemas não circunscritos a um único domínio de conhecimento, onde a heterogeneidade de conceitos e seus significados (heterogeneidade semântica) prevalece. Entenda-se disciplina aqui não só como disciplina científica mas como um domínio de conhecimento cujas características podem também ser determinadas pelos grupos e/ou tecnologias que o geram, por exemplo, conhecimento sobre florestas alimentado por sistemas de geo-processamento versus conhecimento sobre florestas retido por mateiros nativos. Esta definição mais abrangente de disciplina traz consigo a necessidade de resolução de

problemas de heterogeneidade em vários níveis, em especial, heterogeneidade sintática e semântica. Por exemplo, como definir o conceito “floresta” de forma que um sistema de geo-processamento e um sistema sobre ecologia possam ser integrados e possam trocar informações dentro da granularidade e contexto corretos, sem perda de informação? E como estabelecer um relacionamento adequado entre estes diferentes pontos de vista?

Este cenário motiva linhas de pesquisa em computação como Integração de Informação (Information Integration) [42], e Compartilhamento de Conhecimento (Knowledge Sharing) [34]. Ontologias – especificações explícitas de significados de conceitos que podem ser utilizadas por programas e pessoas – potencializam a integração e o compartilhamento de informações, uma vez que formulações de seus significados tornam-se acessíveis e contextualizados com a devida granularidade [41]. Uma abordagem é se ter uma ontologia global reunindo e unificando conceitos que são compartilhados entre os domínios envolvidos [1, 36]. Esta abordagem esbarra na dificuldade de se definir e manter esta ontologia “de consenso” porque, paradoxalmente, tenta amenizar a heterogeneidade inerente aos domínios. Outra abordagem, mais promissora pela maior flexibilidade e manutenibilidade que provê, é cada domínio ter sua ontologia e buscar-se integração através de mapeamentos entre conceitos das diferentes ontologias [30], num estilo peer-to-peer[5]. Aqui, a dificuldade está em se definir mapeamentos que não sejam *ad hoc* entre estas ontologias independentes. Uma forte motivação para este é proporcionar reusabilidade destes mapeamentos entre conceitos.

Esta proposta de dissertação está organizada em três seções além desta introdução. A seção 2 apresenta uma revisão bibliográfica, contendo conceitos básicos relacionados a engenharia de ontologias e operações de integração entre ontologias, como alinhamento e mesclagem. A seção 3 apresenta a proposta deste trabalho, juntamente com suas hipóteses e objetivos, metodologia de desenvolvimento, resultados preliminares e cronograma de execução.

2 Revisão Bibliográfica

Esta seção trata da fundamentação teórica sobre a qual esta proposta se apóia, contextualizando-a em relação a outras abordagens e introduzindo conceitos utilizados. Sowa [33] descreve elegantemente a área na qual esta proposta se insere, explicitando o relacionamento entre lógica e ontologia ¹.

A lógica em si é uma linguagem simples, o nível de detalhe do conhecimento que esta representa depende da escolha dos predicados, o que estritamente falando não pertence à lógica. Os predicados representam uma ontologia de todas as coisas relevantes que existem em relação ao assunto ou aplicação. Diferentes escolhas de predicados representam diferentes compromissos ontológicos. A lógica em si não possui vocabulário para descrever coisas que existem, de forma que a ontologia preenche essa lacuna (p. 15).

Engenharia do Conhecimento é a aplicação da lógica e da ontologia visando a construção de modelos computáveis de algum

¹Ontologia na aspeção da ciência da computação, como veremos adiante.

domínio para algum propósito (p. 132).

2.1 O que é uma Ontologia

Existe uma diferença na definição de ontologia no sentido filosófico e no modo como o termo é usado na Ciência da Computação (CC) [13]. Para a Filosofia, ontologia é o estudo da existência de todas as coisas que formam o mundo, ou seja, a ciência do “*ser*”. Em CC, uma ontologia é vista como um artefato de engenharia que descreve um certo domínio de conhecimento com um vocabulário específico, usando um conjunto de premissas com relação ao sentido pretendido dos termos. Smith [32] sugere uma distinção terminológica entre ontologias baseadas na realidade (*ontologias-R*) e ontologias epistemológicas (*ontologias-E*). *Ontologia-R* é uma ontologia de como todo o universo é organizado e corresponde ao ponto de vista do filósofo. Por sua vez, uma *ontologia-E* se encaixa nos propósitos dos engenheiros de software e cientistas da informação e é definida como uma teoria sobre como um dado indivíduo, grupo, língua ou ciência, conceitua um dado domínio.

É importante salientar que o uso do termo “ontologia” passou por evoluções desde que foi emprestado da filosofia por John McCarthy [17]. Hoje, como assunto de pesquisa, ontologias são importantes em diferentes áreas da CC, tais como: engenharia de conhecimento em Inteligência Artificial, modelagem conceitual em Bancos de Dados e modelagem de domínio em desenvolvimento orientado a objetos. Na CC não existe consenso sobre uma única definição para o termo ontologia. Definições variam conforme o contexto em que são utilizadas e o enfoque aplicado. Contudo, a definição mais citada na área foi cunhada por Gruber: “Uma ontologia é uma especificação explícita de uma conceituação” [38].

2.1.1 Tipos de Ontologias

Existem definições de ontologias pouco restritivas, que consideram simples taxonomias como ontologias; outras, exigem ontologias bem definidas, que fazem uso da lógica na representação de conceitos: “*Ontologias são conjuntos de axiomas lógicos criados para definir o significado pretendido de um vocabulário*” [12]. A comunidade de Representação de Conhecimento [11] distingue as ontologias que são principalmente taxonômicas daquelas que modelam o domínio de uma maneira mais profunda, proporcionando mais restrições à semântica deste. Chamam-se *ontologias leves*, as ontologias que incluem conceitos, taxonomias de conceitos, relações entre conceitos e propriedades que descrevem conceitos. Já as *ontologias pesadas*, além das características das ontologias leves, possuem axiomas e restrições.

Ontologias leves e pesadas podem ser modeladas com diferentes técnicas de modelagem de conhecimento e podem ser implementadas em diversos tipos de linguagens [40]. Dependendo da linguagem em que for expressa, uma ontologia pode ser [11, p. 9]: *altamente informal*, caso expressa em linguagem natural; *semi-informal*, se expressa em uma forma restrita e estruturada da linguagem natural; *semi-formal*, caso expressa em uma linguagem artificial e formalmente definida; e, *rigorosamente formal*, se provê termos meticulosamente definidos com semântica formal, teoremas e, provas de propriedades como *consistência* e *completude*². Adotando terminologia

²Consistência: do inglês *soundness*, termo inspirado na terminologia clássica de lógica de primeira ordem. Dizemos que uma ontologia é consistente (*sound*), se e somente se, não permite

difundida na área utilizaremos nesta proposta o termo *ontologia formal* para nos referirmos às ontologias *semi-formais pesadas*.

2.1.2 Paradigmas de Representação de Ontologias

Dependendo do formalismo adotado para a representação de uma ontologia, é possível restringir as diferentes interpretações dos termos descritos nela. De maneira geral, as abordagens em Inteligência Artificial para representação de ontologias baseiam-se em formalismos como *lógica de primeira ordem*, *frames* e *lógicas descritivas*.

- Lógica de primeira ordem (*First Order Logic-FOL*), lógica com grande poder de expressão, é capaz de definir a semântica de todas outras lógicas, incluindo dela própria [33, p. 41].
- *Frames* [24] são estruturas de dados para representação de situações estereotipadas descritas através de suas características. Permitem a organização do conhecimento visualizada como uma rede de nós e relações.
- Lógicas descritivas (*Description Logics-DL*) são um formalismo [4] cuja teoria é dividida em duas partes distintas: TBox e ABox. Uma TBox (*Terminological Box*) contém conhecimento na forma de uma terminologia e é construída através de declarações que descrevem propriedades gerais dos conceitos. Uma ABox (*Assertional Box*) contém conhecimento assertional que é específico dos indivíduos do domínio de discurso. Em outras palavras, a TBox contém as definições dos conceitos e papéis, enquanto a ABox contém as definições dos indivíduos (instâncias). Desta forma, sistemas em DL permitem a representação de ontologias com três tipos de componentes: conceitos, papéis e indivíduos.

Existem outras possibilidades de representação de ontologias através de ferramentas de áreas como a Engenharia de Software e Banco de Dados, através de diagramas UML e de entidade e relacionamento (ER), respectivamente. Contudo, a semântica destas linguagens só permite a representação de ontologias leves.

2.2 Integração de Ontologias

Com a engenharia de conhecimento gerando ontologias em diferentes organizações e grupos de pesquisa, cada qual buscando soluções para seus próprios problemas e representando conhecimento de domínios específicos, conceitos em comum ou estreitamente relacionados são definidos em várias dessas ontologias.

Esses conceitos relacionados, que na prática podem ter sido representados diferentemente apenas devido à necessidade do domínio e/ou do engenheiro que a especificou, são candidatos ao reuso através da integração dessas ontologias. Contudo, integração de ontologias é uma questão em aberto e, apesar de abordagens recentes [42] proporem soluções adequadas, normalmente estas são aplicáveis apenas em determinadas condições e/ou domínios restritos.

a inferência de conclusões inválidas. Completude: uma ontologia é completa, se e somente se, a partir do vocabulário da ontologia e aplicando as regras de dedução permitidas, permite deduzir todas as conclusões válidas possíveis.

Existem duas abordagens principais para a tentativa de solucionar o problema de integração entre ontologias: alinhamento e mesclagem³ Segundo Noy e Musen [27], podemos definir *alinhamento de ontologias* como o estabelecimento de diferentes tipos de mapeamentos (ou ligações) entre duas ontologias, preservando as ontologias iniciais; já a *mesclagem de ontologias* propõe a geração de uma única ontologia a partir das duas ontologias originais. Ambas as abordagens para integração de ontologias compartilham a mesma necessidade e passo inicial que é a identificação de correspondências entre os conceitos das ontologias envolvidas.

2.2.1 Mapeamentos entre Conceitos

Identificamos cinco principais abordagens para mapeamentos entre conceitos em ontologias: *mapeamentos definidos*, *relações léxicas*, *fundamentação em alto-nível*, *correspondências semânticas* e *métodos probabilísticos e aproximativos*.

- **Mapeamentos definidos:** abordagem mais simples, permite ao engenheiro definir mapeamentos diretos entre conceitos das ontologias envolvidas [30]. Diferentes tipos de mapeamentos entre conceitos podem ser aplicados nesta abordagem, variando de mapeamentos simples um para um a mapeamentos compostos um para vários. Apesar de prover grande flexibilidade, esta abordagem não garante a preservação da semântica, pois o engenheiro é livre para definir mapeamentos arbitrários mesmo que estes não possuam interpretação válida ou produzam conflitos.
- **Relações léxicas:** na tentativa de prover ao menos uma semântica intuitiva para os mapeamentos entre conceitos, esta abordagem estende o modelo da *description logics* através da quantificação das relações entre ontologias emprestadas da linguística [19, 20]. As relações comumente usadas nessas abordagens são *sinonímia*, *hiperonímia*, *hiponímia*, *sobreposição*, *cobertura* e *disjunção*. Apesar dessas relações serem similares às construções usadas em *description logics* elas não possuem uma semântica formal. Consequentemente, possíveis inferências são mais heurísticas do que formalmente fundamentadas.
- **Fundamentação em alto-nível:** prega que para evitar perda de semântica, deve-se permanecer dentro da linguagem formal de representação quando definindo mapeamentos entre diferentes ontologias. Uma maneira direta de seguir esse princípio é relacionar todas as ontologias usadas a uma única ontologia de alto-nível [15]. Pode ser utilizada para resolver alguns casos de conflitos e ambigüidades [14], porém não estabelece correspondências diretas, pois apenas define conexões em termos de super-classes comuns.
- **Correspondências semânticas:** uma abordagem que visa superar ambigüidades decorrentes do mapeamento indireto de conceitos via fundamentação em alto-nível é a tentativa de identificar correspondências semânticas bem fundamentadas entre conceitos de diferentes ontologias. Contudo, de forma a evitar mapeamentos arbitrários entre conceitos, estas abordagens baseiam-se em um vocabulário comum para definição de conceitos entre diferentes ontologias. Wache [43] usa

³Do inglês *alignment* e *merging*, respectivamente.

rótulos semânticos para computar correspondências, já Stuckenschmidt [35] utiliza modelos em *description logics* das diferentes ontologias e mostra que algoritmos de inferência podem ser usados para estabelecer relações entre diferente terminologias.

- **Métodos probabilísticos e aproximativos:** utilizam métodos não-simbólicos para definir mapeamentos entre conceitos. Abordagens como [18, 29] utilizam técnicas de extração em textos como fonte de informação explícita associada a conceitos e usam classificadores para obter pontuações. As pontuações retornadas pelos classificadores de textos são tratadas como probabilidades que auxiliam na determinação dos mapeamentos entre conceitos.

2.2.2 Mesclagem de Ontologias

Propõe resolver o problema de integração entre ontologias através do tratamento das ambigüidades e inconsistências existentes entre as diferentes representações, gerando uma nova ontologia que possua os conceitos descritos nas ontologias originais. Contitui-se em um desafio ainda maior que o mapeamento entre conceitos, visto que eventuais diferenças de granularidade entre as ontologias não podem ser resolvidas como nos mapeamentos que aceitam relações compostas de conceitos. As diferenças terminológicas entre conceitos também devem ser resolvidas adotando-se um vocabulário comum. Abordagens como a Chimaera ⁴, FCA-Merge [37], GLUE [9] e OBSERVER [20], são exemplos de ferramentas que podem ser utilizadas pelo engenheiro em processos interativos de mesclagem de ontologias.

3 Proposta de Trabalho

Nesta seção são apresentadas as principais questões de pesquisa e as hipóteses sobre as quais a proposta se baseia. São explicitados os objetivos e a metodologia que pretendemos utilizar para atingí-los, bem como o cronograma com a duração esperada das etapas. Também são relatadas as decisões de projeto tomadas até o momento e um exemplo ilustrativo do problema delineado. Como as *ontologias formais* (conforme definição da seção 2.1.1) fornecem mais elementos para restrição das possíveis interpretações dos termos, optamos por direcionar a proposta de trabalho a esses tipos de ontologias.

3.1 Questões de Pesquisa

No contexto apresentado na introdução e na revisão de outras abordagens de integração entre ontologias, algumas questões se apresentam:

- Quais características dos componentes de uma ontologia podem ser utilizadas para identificar mapeamentos entre conceitos em diferentes ontologias?
- Apenas a análise local de conceitos, relações entre conceitos e propriedades é suficiente para identificar mapeamentos entre conceitos em diferentes ontologias?

⁴<http://www-ksl.stanford.edu/software/chimaera>

- Os mapeamentos estabelecidos nas abordagens atuais conseguem representar todas as correspondências existentes entre conceitos em diferentes ontologias?
- Que técnicas e/ou métodos podem permitir que instâncias de padrões de mapeamentos entre conceitos sejam relaxados?

A partir das questões apresentadas, tem-se as seguintes hipóteses:

- **Hipótese 1:** é possível identificar e definir *padrões de mapeamento* entre conceitos em ontologias de diferentes domínios. Esses padrões poderiam ser utilizados (instanciados) em mapeamentos entre ontologias de domínios específicos.
- **Hipótese 2:** há mais homogeneidade semântica nas descrições dos significados dos conceitos do que em seus nomes. O corpo das definições dos conceitos (propriedades, axiomas e restrições) e suas relações (com outros conceitos) provêem elementos para comparação com outra ontologia gradativamente mais independentes do domínio à medida que os níveis de definição se aprofundam.
- **Hipótese 3:** assim como o corpo da definição de um conceito e suas relações pode ajudar na identificação de correspondências com os padrões de mapeamentos, acredita-se que as definições dos conceitos relacionados também podem fornecer indicações importantes.

3.2 Objetivos

A presente proposta tem por objetivo desenvolver uma técnica que estabeleça mapeamentos entre conceitos em ontologias distintas, baseados em instâncias de padrões de mapeamentos, através da investigação semântica dos elementos de suas estruturas locais e não-locais. Os contextos local e não-local [25] são elementos da estrutura da ontologia (axiomas, propriedades, outros conceitos, etc.) associados a um conceito de forma direta e indireta, respectivamente. O desenvolvimento da técnica é motivado pela necessidade de suporte computacional à integração de conhecimento em aplicações de enfoque multidisciplinar.

Como objetivos específicos podemos citar:

1. Definir padrões de mapeamento entre conceitos em ontologias de diferentes domínios. Os padrões sugeridos devem utilizar heurísticas para formalizações das relações léxicas entre conceitos (ver seção 2.2.1) [31]. Um ponto importante é a identificação ou aplicabilidade desses padrões entre ontologias de diferentes interpretações de domínio, níveis de abstração e/ou granularidade.
2. Investigar abordagens para relaxar os mapeamentos definidos. Um estudo multidisciplinar pode ser enriquecido por uma equivalência ou contradição entre conceitos, mas também, e talvez mais frequentemente, por relações menos restritas como complementação ou similaridade. Para tanto, serão consideradas abordagens aproximativas para representação de conhecimento e inferência, como satisfação e propagação de restrições (*constraints*) [41, 21].
3. Implementar um protótipo funcional da técnica desenvolvida. Será investigada a possibilidade de se implementar o protótipo como um

plug-in de um ambiente para modelagem de ontologias, proporcionando assim o aproveitamento das funcionalidades disponíveis no ambiente. A inclusão do protótipo como recurso de uma ferramenta potencializaria a utilização deste pela comunidade que a adota.

3.3 Metodologia

As etapas a serem cumpridas no trabalho proposto são:

1. levantamento bibliográfico acerca dos formalismos e linguagens existentes de representação de ontologias e mapeamento entre conceitos.
2. caracterização de relações entre conceitos interdisciplinares.
3. estudo dos métodos clássicos e flexíveis de satisfação de restrições.
4. definição de um estudo de caso.
5. identificação de padrões de correspondências entre conceitos em ontologias de diferentes domínios;
6. definição de padrões de mapeamento entre conceitos.
7. experimentação de métodos flexíveis de satisfação de restrições para comparação da descrição dos conceitos com os padrões de mapeamentos.
8. desenvolvimento de método de instanciação dos padrões de mapeamentos, eventualmente relaxados, para relacionar conceitos entre diferentes ontologias.
9. implementação do protótipo.
10. análise da qualidade dos resultados segundo um framework de avaliação e comparação de métodos de mapeamentos.

A comparação dos formalismos e linguagens de representação de ontologias (etapa 1) visa identificar quais deles fornecem elementos suficientes para possibilitar a análise proposta dos mapeamentos entre conceitos de diferentes ontologias, como também quais provêem poder de expressão para a definição de padrões de mapeamentos.

Para caracterizar as relações entre conceitos interdisciplinares (etapa 2), pretende-se analisar fontes da Linguística e Ciências Cognitivas. Projetos como o WordNet [23] fornecem caracterizações léxicas das relações entre conceitos. Na área de Inteligência Artificial, abordagens terminológicas de mapeamentos [19] também utilizam relações léxicas para mapeamentos entre conceitos.

A análise dos métodos clássicos e flexíveis de satisfação de restrições [8, 22] (etapa 3) visa identificar a aplicação destes ao mecanismo de estabelecimento de mapeamentos relaxados na instanciação dos padrões de mapeamento.

A definição de um estudo de caso (etapa 4) objetiva fornecer um cenário concreto para a identificação dos padrões de mapeamentos entre conceitos. São necessárias duas ontologias descritas no formalismo e linguagem definidos na etapa 1. A *Ecolingua* [6], ontologia que descreve quantidades no domínio de Ecologia, foi escolhida por sua temática ambiental que é parte do projeto de pesquisa MOeMA-IS (Metadados, Ontologias e Indicadores de Sustentabilidade Integrados a Modelagem Ambiental⁵). Para a montagem do

⁵O projeto MOeMA-IS é financiado pela FAPEAM e integrado por pesquisadores da UFAM, INPA, CEFET e SUFRAMA.

cenário é necessária uma segunda ontologia, a qual está sendo desenvolvida como parte do trabalho proposto, também no contexto do projeto MOeMA-IS. Esta ontologia aborda a temática econômica dos indicadores de sustentabilidade segundo proposta das Nações Unidas [39]. As ontologias escolhidas auxiliarão a identificação de padrões de mapeamentos entre conceitos em ontologias de diferentes domínios (etapa 5).

A partir das relações entre conceitos identificadas na etapa 2, deve-se buscar na literatura subsídios para a formalização de padrões de mapeamento entre conceitos (etapa 6). Trabalhos como [3], por exemplo, fornecem diferentes graus de formalização para a relação *parte-todo*, apresentando inclusive os problemas relacionados às formalizações propostas.

A experimentação de métodos flexíveis de satisfação de restrições na comparação de axiomas com os padrões de mapeamento (etapa 7) e o desenvolvimento do método de instanciação (etapa 8) destes serão utilizados no desenvolvimento do protótipo (etapa 9) que fornecerá dados para a análise qualitativa dos mapeamentos que serão avaliados provavelmente segundo o framework proposto por [26] (etapa 10).

3.4 Resultados Preliminares

Após análise dos formalismos de representação de ontologias (1) optamos por uma abordagem baseada em *frames* e *Description Logics (DL)*. O formalismo baseado em frames proporciona uma forma de organização do conhecimento aceita como a mais próxima da representação mental dos humanos, possuindo similaridades com o paradigma de representação orientado a objetos. A utilização de uma sintaxe baseada no paradigma de frames proporciona legibilidade e facilidade geral de uso em uma linguagem para especificação de ontologias. Os frames agrupam informações sobre cada classe facilitando a leitura e o entendimento, particularmente para os usuários não familiarizados com DL. Já o formalismo em DL possibilita a utilização de mecanismos de inferência, tais como checagem de restrições e classificadores automáticos.

A partir da escolha dos formalismos, buscou-se uma linguagem para descrição de ontologias que os adotassem. De acordo com a motivação exposta na introdução, a técnica de mapeamento deve suportar ontologias multidisciplinares desenvolvidas para anotação semântica na Web. Portanto, tomou-se por base as características que o W3C identificou como desejáveis para uma linguagem Web para descrição de ontologias, a saber: facilidade de uso e entendimento, basear-se em paradigmas populares de representação de conhecimento, ser formalmente especificada, possuir poder expressivo adequado, prover possibilidade de suporte a inferência automática e estender padrões web existentes (XML, RDF e RDF schema, etc). A partir dessas características e das necessidades do trabalho, identificou-se a OWL (Ontology Web Language) como linguagem de modelagem. A especificação da OWL contempla 3 (três) sub-linguagens: OWL Full, OWL-DL e OWL Lite, das quais a OWL-DL foi escolhida devido ao seu poder de expressão aliado à sua decidibilidade, características herdadas da família *SHOIN(D)* das *description logics*, que por sua vez é uma extensão da influente *SHOQ(D)* [16]. A especificação formal da OWL foi influenciada pela DL, enquanto sua estrutura superficial foi influenciada pelo paradigma de frames. Já o RDF/XML influenciaram sua sintaxe, como um requisito de compatibilidade com o RDF.

Atualmente poucas ferramentas suportam a modelagem de ontologias em OWL, as principais são: WebODE [2] e Protégé-2000 [28]. A Protégé-2000 foi escolhida por ser de código aberto (Mozilla license) e possuir arquitetura extensível, possibilitando o desenvolvimento de plug-ins. O próprio suporte à linguagem OWL no Protégé é provido através de um plug-in. A comunidade em torno deste plugin-OWL é bastante ativa e trabalha na evolução da ferramenta para atender a especificação completa da OWL [7].

As relações caracterizadas pela linguística (2) no projeto WordNet serviram de ponto de partida para alguns tipos de padrões de mapeamentos que podem ser utilizados:

- **Sinonímia:** relação de similaridade de significado entre palavras, duas expressões são sinônimas se a substituição de uma por outra nunca muda o valor verdade de uma sentença na qual a substituição foi feita. Uma definição *relaxada* da sinonímia leva em consideração o contexto da sentença: duas expressões são sinônimas em um contexto linguístico se a substituição de um termo por outro no contexto não altera o valor verdade. As definições de sinonímia baseadas em valor-verdade a tornam um conceito discreto, enquanto alguns filósofos e psicólogos concordam que a sinonímia é melhor definida como o final de um intervalo contínuo ao longo do qual a similaridade de significado pode ser determinada.

$$Quantity@Ecolingua \equiv Quantity@IS-E$$

,conceito *Quantity* na ontologia *Ecolingua* é *sinônimo* do conceito *Quantity* na ontologia *IS-E*⁶.

- **Antonímia:** na linguística é uma relação que exige cautela na sua identificação pois significados de palavras podem ser conceitualmente opostos, mas não serem antônimos. Tal qual a sinonímia, a relação entre duas palavras antônimas não é simétrica. Uma palavra ser X, nem sempre implica que seu antônimo é não-X.

$$Non-Physical-Qty@Ecolingua \perp PhysicalQuantity@IS-E$$

- **Hiperonímia e hiponímia:** são relações de subordinação entre significados de palavras, significam “*subclasse de*” ou “*é um*”.

$$LengthMeasure@IS-E \sqsubseteq DimensionUnit@Ecolingua$$

- **Meronímia e holonímia:** são conceitos *parte-todo* e significam respectivamente as relações “*parte de*” e “*possui um*”. A relação meronímica é transitiva (com qualificações) e assimétrica, podendo ser utilizada para construir hierarquia de partes (com algumas reservas, pois um merônimo pode possuir varios holônimos).

$$Amount-of-Time@Ecolingua \succeq TimeMeasure@IS-E$$

Quanto ao estudo de caso (etapa 4), uma ontologia de indicadores de sustentabilidade com enfoque em economia foi parcialmente especificada, enquanto a *Ecolingua* está sendo portada para OWL-DL por outro integrante

⁶No apêndice constam as hierarquias de ambas ontologias.

do projeto MOeMA-IS. Esta experiência de especificação de uma ontologia deixou claro que a dificuldade no estabelecimento de padrões de mapeamento entre conceitos de ontologias em diferentes domínios (etapa 5) advém da diferença de granularidade de conceitos e da diversidade de especificações em que um mesmo conceito pode ser expresso. Cada domínio possui suas peculiaridades e as relações definidas entre seus conceitos varia dependendo de quanto utilizado é esse conceito na definição de outros. Portanto, ontologias densas, possuindo muitas relações entre conceitos e restringindo fortemente o significado de cada um deles, possivelmente apresentarão poucos mapeamentos com ontologias superficiais de poucas relações entre conceitos. O desempenho da técnica de mapeamentos baseada em padrões será dependente desses fatores.

3.5 Cronograma

O desenvolvimento deste trabalho contempla as etapas da metodologia descritas na seção 3.3 e seguirá o cronograma a seguir:

Etapas da Tese	Dez '04	Jan '05	Fev '05	Mar '05	Abr '05	Mai '05	Jun '05	Jul '05	Ago '05	Set '05	Out '05	Nov '05	Dez '05	Jan '06	Fev '06
Etapa 1	■														
Etapa 2	■	■													
Etapa 3		■	■	■											
Etapa 4		■	■	■											
Etapa 5			■	■	■	■	■								
Etapa 6				■	■	■	■								
Etapa 7					■	■	■	■	■	■					
Etapa 8						■	■	■	■	■	■				
Etapa 9						■	■	■	■	■	■				
Etapa 10											■	■	■		
Dissertação				■	■	■	■	■	■	■	■	■	■	■	■

Tabela 1: Cronograma de trabalho.

Referências

- [1] Y. Arens, C.N. Hsu, and C.A Knoblock. Query processing in the SIMS information mediator. *Advanced Planning Technology*, 1996.
- [2] J. Arpírez, O. Corcho, M. Fernández-López, and A. Gómez-Pérez. Webode: a scalable ontological engineering workbench. In *In Proceedings of the First International Conference on Knowledge Capture*, pages 6–13, Victoria, Canada, 2001.
- [3] Alessandro Artale, Enrino Franconi, and Nicola Guarino. Open problems with part-whole relations.

- [4] Franz Baader, Ian Horrocks, and Ulrike Sattler. Description logics as ontology languages for the semantic web. In *Proceeding of the International Workshop on Ontologies*, 2002.
- [5] P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. Contextualizing ontologies. *Journal of Web Semantics*, 2004.
- [6] Virgínia Brilhante. An ontology for quantities in ecology. In *Proceedings of the Brazilian Symposium on Artificial Intelligence*, São Luiz, Maranhão, Brasil, 2004.
- [7] M. Dean and G. Schreiber. Owl web ontology language referencer. Technical report, W3C, 2003. <http://www.w3.org/TR/owl-ref/>.
- [8] Rina Dechter. *Constrain Processing*. Morgan Kaufmann, 2003.
- [9] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In *Proceedings of the 11th International World Wide Web Conference (WWW 2002)*, Honolulu, Hawaii, 2002.
- [10] Frederico Fonseca, James Martin, and Andre Rodriguez. From geo to eco-ontologies.
- [11] Asunción Gómez-Perez, Mariano Fernández-López, and Oscar Corcho. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer-Verlag London, 2004.
- [12] N Guarino and P Giaretta. Ontologies and knowledge bases: towards a terminological clarification. In Mars N, editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS'95)*, pages 25–32. IOS Press, Amsterdam, The Netherlands, University of Twente, Enschede, The Netherlands,, 1995.
- [13] Nicola Guarino. *Formal Ontology and Information Systems*. IOS Press, Amsterdam, Netherlands, 1998.
- [14] Jeff Heflin and James Hendler. Extreme markup languages 2000, 2000.
- [15] Jeff Heflin, James Hendler, and Sean Luke. Shoe: A knowledge representation language for internet applications. Technical Report Technical CS-TR-4078, Institute for Advanced Computer Studies, University of Maryland, 1999.
- [16] Ian Horrocks and Ulrike Sattler. Ontology reasoning in the *SHOQ(d)* description logic. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI2001)*, pages 199–204, 2001.
- [17] McCarthy J. Circumscription - a form of non-monotonic reasoning. *Art. Int.* 5, 13, pages 27–39, 1980.
- [18] Martin S. Lacher and Georg Groh. Facilitating the exchange of explicit knowledge through ontology mappings. *American Association for Artificial Intelligence*, 2001.
- [19] Andreia Malucelli and Engénio Oliveira. Towards to similarity identification to help in the agents' negotiation. In *Proceeding of the International Workshop on Ontologies*, 2004.
- [20] E. Mena, V. Kashyap, A. Sheth, and A. Illarramendi. Observer: An approach for query processing in global information systems based on

- interoperability between pre-existing ontologies. In *Proceedings 1st IFCIS International Conference on Cooperative Information Systems (CoopIS'96)*, Brussels, 1996.
- [21] I. Miguel and Q. Shen. Solution techniques for constraint satisfaction: Advanced approaches. *Artificial Intelligence Review* 15(4), pages 269–293, 2001.
 - [22] Ian Miguel and Qiang Shen. Hard, flexible and dynamic constraint satisfaction. School of Artificial Intelligence - University of Edinburgh.
 - [23] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. Introduction to wordnet:an online lexical database. *International Journal of Lexicography* 3(4), pages 235–244, 1990.
 - [24] Marvin Minsky. A framework for representing knowledge. In P. Winston, editor, *The Psychology of Computer Vision*, pages 221–280. McGraw-Hill, New York, 1975.
 - [25] Natalya Noy and Mark Musen. Anchor-prompt: Using non-local context for semantic matching. In A. Gómez-Pérez, M. Grüninger, H. Stuckenschidt, and M. Uschold, editors, *Proceedings of IJCAI'01 Workshop on Ontologies and Information Sharing*, pages 63–70, Seattle, Washington, 2001.
 - [26] Natalya Noy and Mark Musen. Evaluating ontology-mapping tools: Requirements and experience. 2003.
 - [27] Natalya Noy and Mark A Musen. Smart: Automated support for ontology merging and alignment. In Musen Gaines, Kremer, editor, *12th Banff Workshop on Knowledge Acquisition, Modeling and Management*, pages 4–7:1–20, Banff, Alberta, Canada, 1999.
 - [28] NF. Noy, RW. Fergerson, and MA. Musen. The knowledge model of protégé-2000: Combining interoperability and flexibility. In *12th International Conference in Knowledge Engineering and Knowledge Management EKAW0*, pages 17–32, Juan-Les-Pins, Grance. Springer-Verlag, Berlin, Germany.
 - [29] Sushama Prasad, Yun Peng, and Timothy Finin. Using explicit information to map between two ontologies. 2002.
 - [30] A.D. Preece, H.J. Hui, W.A. Gray, P. Marti, T.J.M. Bench-Capon, D.M. Jones, and Z. Cui. The kraft architecture for knowledge fusion and transformation. In Springer, editor, *Proceedings of the 19th SGES International Conference on Knowledge-Based Systems and Applied Artificial Intelligence(ES99)*, 1999.
 - [31] David Robertson and Jaume Agustí. *Software Blueprints*. Addison Wesley, ACM Press, 1999.
 - [32] B. Smith. An introduction to ontology. *THE Ontology of Fields*, 1998.
 - [33] J. F. Sowa. *Knowledge Representation - Logical, Philosophical and Computational Foundations*. Brooks/Cole, Croton-on-Hudson, New York, 2000.
 - [34] H. Stuckenschmidt and F. van Harmelen. *Information Sharing on the Semantic Web*. Springer, 2004.

- [35] H. Stuckenschmidt, H. Wache, Ta. Vogele, and U. Visser. Enabling technologies for interoperability. In U. Visser and H. Pundt, editors, *Proceedings of the Workshop of the 14th International Symposium of Computer Science for Environmental Protection*, pages 35–46, Bonn, Germany, 2000.
- [36] Heiner Stuckenschmidt, Frank van Harmelen, Dieter Fensel, Michel Klein, and Ian Horrocks. Catalogue integration: A case study in ontology-based semantic translation. Technical Report IR-474, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, 2000.
- [37] G. Stumme and A. Maedche. Fca-merge: Bottom-up merging of ontologies. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI 2001)*, pages 225–234, Seattle, Washington, 2001. Morgan Kaufmann.
- [38] Gruber TR. A translation approach to portable ontology specification. *Knowledge Acquisition* 5(2), pages 199–220, 1993.
- [39] United Nations. *Indicators of Sustainable Development: Guidelines and Methodologies*, joanne disano edition, 2003.
- [40] Uschold and Grüninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review* 11(2), pages 93–155, 1996.
- [41] M. Uschold and M. Grüninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), pages 93–155, 1996.
- [42] H. Wache, Ta. Vogele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hubner. Ontology-based integration of information - a survey of existing approaches. In *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing*, pages 108–117, 2001.
- [43] Holger Wache. Towards rule-based context transformation in mediators. In Infix-Verlag, editor, *International Workshop on Engineering Federated Information Systems (EFIS'99)*, Khlungsborn, Germany.

A Ontologias

A.1 Ecolingua

Figura 1: Hierarquia da *Ecolingua*

A.2 Indicadores de Sustentabilidade - Temática Econômica

Figura 2: Hierarquia da IS-E

Figura 3: Hierarquia da IS-E - Estrutura Econonômica

Figura 4: Hierarquia da IS-E - Indicadores de Consumo e Padrões de Produção

