

### 3 Trabalhos Relacionados

O desenvolvimento de um Processo Unificado nos moldes do KUP exige uma análise aprofundada de diversos tipos de trabalhos relacionados. Nesta dissertação serão descritas quatro categorias de trabalhos, sendo: (i) metodologias, métodos e processos para engenharia de ontologias; (ii) projetos de gestão de conhecimento e portais semânticos que utilizem ontologias e a web semântica e (iii) trabalhos envolvendo a integração de ontologias e (iv) abordagens utilizando ontologias para gestão distribuída de conhecimento.

#### 3.1. Metodologias, métodos e processos para engenharia de ontologias

Diversos grupos utilizam abordagens próprias para engenharia de ontologias, (Gómez-Pérez, 1999), muitas delas desenvolvidas até mesmo fora do contexto da Web Semântica. Algumas dessas metodologias, métodos e processos para projeto e desenvolvimento de ontologias são descritas a seguir.

##### 3.1.1. METHONTOLOGY

Methontology (Fernández-López et al., 1997) é uma metodologia que define o processo de criação de ontologias desde o seu início, passando pelas etapas de planejamento, especificação, aquisição de conhecimento conceitualização, formalização, integração, implementação, avaliação, documentação e manutenção. É a mais completa e abrangente das metodologias estudadas, mas não aborda diversas atividades fundamentais de um processo de engenharia de ontologias, as quais serão apresentadas no capítulo 5. É baseada na idéia de prototipação e evolução como a abordagem mais adequada para o ciclo de vida de uma ontologia e dá uma especial ênfase ao reuso das mesmas. Suas atividades podem ser divididas em:

1. Atividades de Gerência de Projeto
2. Atividades Orientadas ao Desenvolvimento
3. Atividades de Suporte

O ambiente para construção de ontologias que utiliza este framework é chamado ODE (*Ontology Design Environment*) (Blázquez et al., 1998). Seu objetivo é dar suporte aos engenheiros de ontologia durante o ciclo de vida do processo de desenvolvimento. Este ambiente procura automatizar cada atividade de desenvolvimento da ontologia e a integração dos resultados de cada fase com as entradas da fase seguinte.

### **3.1.2. Metodologia On-To-Knowledge**

O projeto On-To-Knowledge (Fensel et al., 2002),(Sure et al., 2002b) tem como principal objetivo permitir a aplicação de ontologias a informações disponibilizadas eletronicamente na web, com o intuito de aprimorar os processos de Gestão de Conhecimento em grandes organizações. Entre suas contribuições podem ser citados um extenso suporte a aquisição, manutenção e acesso a grandes volumes de dados semi-estruturados e textuais e uma metodologia que provê orientações para a introdução de conceitos e ferramentas de Gestão de Conhecimento em empresas. Essa metodologia aborda os objetivos que precisam ser alcançados por esse tipo de ferramenta e é baseada em uma análise dos processos de negócios e papéis que os trabalhadores intensivos em conhecimento têm no contexto da organização.

Para a realização de projetos nesse sentido propõe uma série de atividades, listadas a seguir:

1. Meta-processo de Conhecimento: são realizadas atividades para introdução de uma nova solução de gestão e manutenção do conhecimento numa organização.
  - 1.1. Estudo de viabilidade
  - 1.2. Inicialização
  - 1.3. Refinamento
  - 1.4. Avaliação
  - 1.5. Aplicação e evolução
  
2. Processo de Conhecimento: O foco deste processo é a utilização contínua da aplicação de gestão de conhecimento implementada na organização. Neste processo são realizadas as seguintes atividades:
  - 2.1. Criação do conhecimento e/ou importação de documentos e metadados

- 2.2. Captura do conhecimento
- 2.3. Recuperação e acesso ao conhecimento
- 2.4. Utilização do conhecimento

### **3.1.3. Metodologia de Uschold & King**

Uschold e King elaboraram esta metodologia no contexto do desenvolvimento da *Enterprise Ontology* (Uschold and King, 1995), uma ontologia para dar suporte à modelagem de processos empresariais. Esta metodologia descreveu a utilização de quatro etapas principais para o desenvolvimento dessa ontologia, as quais podem ser aplicadas no desenvolvimento de outras ontologias:

1. Identificação do propósito e escopo:
2. Construção da ontologia, dividida em três passos:
  - a. *Captura da ontologia*
  - b. *Codificação*
  - c. *Integração de ontologias existentes*
3. Avaliação
4. Documentação

### **3.1.4. Metodologia de Grüninger & Fox**

Esta metodologia também é baseada na experiência de desenvolvimento de uma ontologia, no contexto do projeto TOVE (Grüninger e Fox, 1995). Este projeto descreve o domínio de processos de negócios e a modelagem de atividades com o uso de ontologias, permitindo a transformação de cenários informais expressos em linguagem natural em modelos computáveis, expressos em linguagem formal. As seguintes etapas são propostas nesta metodologia:

1. Captura dos cenários motivacionais
2. Formulação de questões de competência informais
3. Especificação da terminologia da ontologia numa linguagem formal
4. Formulação das questões de competência formais
5. Especificação dos axiomas
6. Verificação da completude da ontologia

### 3.1.5. Método SENSUS

Este método (Swartout, 1996) está baseado na ontologia SENSUS, criada pelo grupo de processamento de linguagem natural do ISI (Information Science Institute). Esta ontologia foi criada para prover uma estrutura conceitual ampla que pudesse ser utilizada no desenvolvimento de tradutores automáticos. A SENSUS contém tanto conceitos de alto nível quanto específicos, o que justifica a denominação de ontologia de ampla cobertura, possuindo aproximadamente 70.000 conceitos organizados hierarquicamente de acordo com seu nível de abstração. No entanto, esta ontologia foca principalmente nos níveis de abstração alto e médio, deixando a especificação de termos mais concretos - de acordo com o domínio em análise desejado - para as ontologias que serão geradas a partir dela.

A construção da ontologia de domínio parte da ontologia SENSUS, com a adição de novos termos específicos e a retirada de termos irrelevantes. A seguir, os passos propostos pelo Método.

1. Um conjunto de termos é selecionado como “semente” para a ontologia.
2. Esses termos são localizados ou incluídos na ontologia SENSUS manualmente.
3. Todos os conceitos no caminho desses termos até a raiz da SENSUS – segundo a hierarquia de conceitos – são incluídos na nova ontologia.
4. Termos relevantes para o domínio são adicionados.
5. Termos centrais, isto é, que possuem muitas relações com outros termos da ontologia podem justificar a introdução de uma sub-árvore completa da SENSUS na ontologia gerada,

O software utilizado para construir ontologias seguindo esta metodologia é o Ontosaurus (Swartout, 1996) Ontosaurus é um servidor que pode ser acessado através da web e representa o conhecimento na linguagem de programação LOOM (Ding, 2000), permitindo tradução para Ontolingua, KIF e C++.

### 3.1.6. Método Cyc

O método Cyc foi desenvolvido a partir da experiência de desenvolvimento da base de conhecimento Cyc (Lenat et al., 1990). Esta base de conhecimento tem como objetivo ser um compêndio do conhecimento humano, contendo uma grande quantidade de conhecimento de senso comum descrevendo objetos e ações do cotidiano. A base Cyc pode ser usada na identificação de inconsistências,

contradições e violações do senso comum, ou mesmo servir como uma ontologia de alto nível, dando suporte à recuperação de informações relacionadas na Web. Pode também ser usada na identificação de inconsistência em bancos de dados, como por exemplo na verificação de informações sobre um empregado que foi demitido antes de nascer (Lenat, 1995).

Este método é dividido em duas fases principais. A cada fase é desenvolvida uma ontologia contendo os conceitos mais abstratos que será então refinada posteriormente, para representação do conhecimento desejado. Na primeira fase deve-se a identificar e codificar o conhecimento explícito e implícito existente nas fontes de conhecimento disponíveis, para que, na segunda fase, novos conceitos e relações sejam inferidas com o auxílio de ferramentas de processamento de linguagem natural e/ou aprendizado de máquina.

### **3.1.7.Método KACTUS**

A idéia principal deste método (Bernaras et al., 1996), desenvolvido no contexto do projeto Espirit Kactus é permitir a construção de uma ontologia que representa o conhecimento necessário a uma aplicação durante o seu desenvolvimento. Esta ontologia de aplicação pode ser construída a partir da reutilização de outras ontologias de domínio e de tarefas (Guarino, 1998) e pode ser integrada às próximas ontologias construídas, promovendo, assim, um intenso reuso de conhecimento.

A cada aplicação desenvolvida são desempenhadas as seguintes tarefas:

1. Especificação da aplicação
2. Projeto preliminar
3. Refinamento e estruturação da ontologia

## **3.2. Gestão de Conhecimento utilizando Ontologias e a Web Semântica: Portais Semânticos**

### **3.2.1.SEAL e SEAL-II**

Um dos mais importantes exemplos de desenvolvimento de frameworks e metodologias para a construção de portais no contexto da Web Semântica é o projeto SEAL (SEmantic portal), atualmente em sua segunda versão – SEAL-II. Os

autores deste trabalho desenvolveram uma abordagem genérica para o projeto de portais semânticos que utiliza os dados semânticos tanto para automatizar tanto o acesso quanto a própria manutenção do portal (Maedche et al., 2001).

Ontologias constituem a base desta abordagem. A origem do SEAL está no projeto Ontobroker (Decker et al., 1999), que foi concebido para dar suporte a buscas semânticas e compartilhamento de conhecimento na Web (Benjamins et al., 1998), (Hocho et al., 2001).

### **3.2.2. KA2**

Outro trabalho relacionado a Gestão de Conhecimento utilizando ontologias e a Web Semântica através da formação de comunidades é o portal KA2 da "Knowledge Annotation initiative of the Knowledge Acquisition community" (Benjamins et al., 1999)). Esta iniciativa foi concebida para permitir a recuperação semântica do conhecimento criado pela comunidade de *Knowledge Acquisition*. Uma ontologia foi criada para estruturar o conhecimento manipulado, através de uma colaboração internacional de pesquisadores. Documentos da comunidade passaram então a ser anotados através desta ontologia para permitir a inferência de conhecimento implícito a partir de fatos descritos na ontologia e as regras de inferência que a compõem. Um portal de demonstração está disponibilizado em <http://ka2portal.aifb.uni-karlsruhe.de>, na opção KBNavigate, assim como a ontologia criada.

A ontologia KA2 consiste de (i) conceitos definindo e estruturando termos importantes no domínio, (ii) seus atributos especificando propriedades e relações e (iii) regras de inferência para a geração de novo conhecimento. A linguagem de representação escolhida foi F-Logic (Kifer et al., 1995). Trechos dessa ontologia podem ser vistos em (Staab et al., 2001).

### **3.2.3. University of Tatung Semantic Portal**

Um Portal Semântico foi também desenvolvido para a biblioteca da Universidade de Tatung, em Taiwan (Yeh et al., 2002). A arquitetura do portal está dividida em duas camadas: uma camada de conteúdo e outra de aplicações. O conteúdo do portal é representado em RDF e a camada de aplicações possui funcionalidades para acesso e indexação, busca conceitual, navegação, fórum e

personalização. São também oferecidos mecanismos para edição de ontologias, *parsing* do RDF e extração de informações.

Segundo os autores deste projeto esta separação entre conteúdo e funcionalidades permite um maior reuso do sistema em outros domínios que não a biblioteca utilizada no estudo de caso.

### 3.3.Integração de ontologias

Ultimamente, diversas ontologias vêm sendo desenvolvidas para os mais diferentes fins. É natural – pelos mais diferentes motivos - que grupos de pesquisadores responsáveis por essas ontologias precisem trabalhar em conjunto, o que torna necessário desfazer as disparidades entre as mesmas, quando estas descrevem um mesmo domínio ou domínios com uma significativa interseção entre si. Neste sentido, duas abordagens podem ser utilizadas: A primeira consiste na geração de uma ontologia que represente a integração (fusão) das ontologias díspares. A segunda, chamada de alinhamento, é realizada através do estabelecimento de ligações entre as ontologias para que uma possa passar a usar as informações descritas pela outra (Noy et al., 1999).

A Figura 8 ilustra as diferenças entre integração e alinhamento de ontologias. Sendo O1 e O2 as ontologias originais, as linhas pontilhadas na figura representam as transições sofridas no processo, e a área cinza o resultado do mesmo. O resultado da integração é uma única ontologia, O. No caso de alinhamento, por outro lado, as duas ontologias permanecem, agora com relacionamentos entre si, representados pelas linhas preenchidas. Geralmente, esta última abordagem é aplicada quando os domínios descritos são complementares entre si.

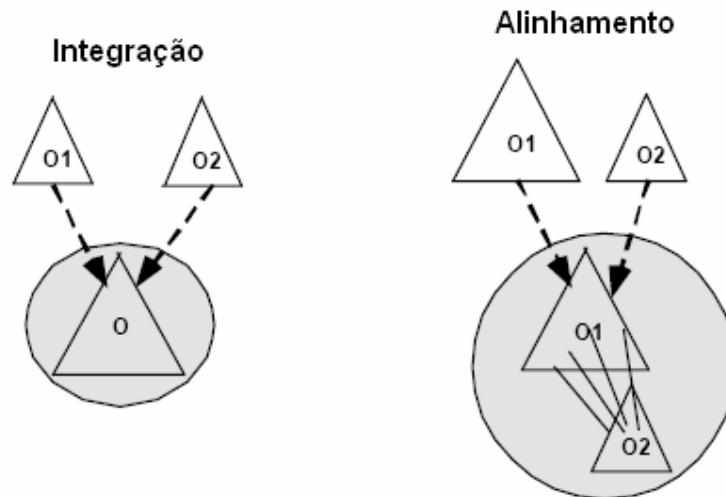


Figura 8 - Integração e Alinhamento de Ontologias

Nesta categoria de trabalhos relacionados serão descritos o PROMPT e o Chimaera/Ontolingua.

### 3.3.1. PROMPT

PROMPT é um algoritmo que implementa uma abordagem semi-automática para prover a integração e alinhamento de ontologias. O PROMPT recebe como entrada duas ontologias e guia o usuário na criação de uma ontologia integrada. Os passos desse algoritmo são descritos a seguir:

O primeiro passo do algoritmo consiste na criação de uma lista de proximidades baseada nos nomes dos conceitos das duas ontologias de entrada. O usuário então pode ativar uma operação baseada na lista de sugestões apresentadas pelo PROMPT ou usando um ambiente de edição de ontologias para especificar a operação desejada diretamente. O PROMPT executa a operação e atualiza a lista de sugestões baseado na ontologia resultante. Esta nova lista pode identificar conflitos causados pela operação anterior e apontar possíveis soluções. O ciclo então se repete até que todas as modificações necessárias tenham sido feitas e todos os conflitos eliminados. Diversos mecanismos para calcular a similaridade entre nomes de conceitos podem ser usados, como identificado em (Noy et al., 2000).



Os autores de PROMPT oferecem ainda uma implementação do algoritmo baseada no Protégé que permite a utilização de diferentes mecanismos de comparação de similaridade lingüística. Esta implementação concentra-se na geração de sugestões baseadas na estrutura da ontologia e nas ações dos usuários. Um resumo da abordagem é apresentado a seguir:

Foi identificado um conjunto de operações típicas para integração e alinhamento de ontologias. Para cada uma dessas operações foram associadas: (a) mudanças que podem ser automaticamente efetuadas pelo PROMPT, (b) sugestões que podem ser apresentadas ao usuário e (c) conflitos que podem ser causados por estas mudanças e que o usuário precisa decidir como proceder. A cada operação executada esses diferentes conjuntos são atualizados.

Esse conjunto operações possíveis foi gerado com base nas operações normalmente realizadas durante a edição de ontologias, somadas a operações de alinhamento e integração, como:

- fusão de classes,
- fusão de propriedades,
- fusão de ligações entre classes e propriedades,
- *deep-copy* - cópia de uma classe de uma ontologia para outra (incluindo todo o ramo da hierarquia partindo da classe raiz até esta classe e seus relacionamentos)
- *shallow-copy* - cópia de uma classe de uma ontologia para outra (apenas a classe em questão)

Com isso, foram identificados os seguintes conflitos que podem ocorrer:

- conflitos de nomes – conceitos com mesmo nome e semânticas distintas,
- referências quebradas – conceitos que possuem relações com conceitos ou propriedades que não existem,
- redundâncias na hierarquia de classes,
- valores de propriedades que violam relações de herança/generalização.

Maiores detalhes sobre o algoritmo PROMPT e a sua implementação na ferramenta Protege (Grosso et al. 1999) podem ser encontradas em (Noy et al., 2000).

### **3.3.2. Chimaera/Ontolingua**

Chimaera (McGuinees et al., 2000) é um ambiente de edição, integração e diagnóstico de ontologias, totalmente baseado na web. Foi desenvolvido com o objetivo de solucionar os problemas relacionados à representação de conhecimento e raciocínio (*reasoning*) na Web, além de dar suporte às tarefas de criação e manutenção de ontologias. O Chimaera foi construído sobre o Ontolingua (Farquhar et al., 1997), que provê um ambiente colaborativo e distribuído para navegação, criação, edição e utilização de ontologias. O ambiente possui uma biblioteca de ontologias modulares e reutilizáveis, que devem ser combinadas e estendidas para a criação de novas ontologias através da ferramenta de autoria disponibilizada.

## **3.4. Gestão Distribuída de Conhecimento utilizando Ontologias**

### **3.4.1. OntoShare**

OntoShare (Davies et al., 2003) é um ambiente na Web para compartilhamento de conhecimento baseado em ontologias que permite a modelagem dos interesses de cada usuário na forma de um perfil individual. Esses perfis correspondem a um conjunto de tópicos ou conceitos da ontologia definida em RDF Schema para os quais o usuário identifica certo interesse. O OntoShare sumariza e extrai palavras-chave de fontes de informação – que podem ser páginas na web ou documentos em geral compartilhados por algum usuário da comunidade – e disponibiliza essas informações a outros usuários com um perfil que demonstre potencial interesse no conhecimento extraído.

O OntoShare é usado para armazenar, recuperar, sumarizar e notificar usuários sobre determinada informação identificada como valiosa, compartilhada por algum usuário da comunidade. Este compartilhamento é feito através de um cliente Java, a partir do qual o usuário pode fornecer meta-dados sobre a informação que está sendo disponibilizada. Neste ponto o sistema irá comparar este conteúdo com os conceitos da ontologia da comunidade. Essa comparação é feita através de um algoritmo que calcula a proximidade entre vetores de termos (Granovetter, 1982). O

sistema sugere então ao usuário um conjunto de conceitos que podem ser relacionados ao conteúdo sendo submetido. O usuário pode então aceitar as recomendações do sistema ou modificá-las sugerindo alternativas ou conceitos adicionais que podem ser associados ao conteúdo sendo submetido. Em seguida, o OntoShare cria um resumo desse conteúdo através de uma ferramenta (ViewSum) que extrai sentenças-chave do texto, baseando-se na frequência de palavras e frases. Esta técnica está fundamentada na análise de coesão léxica (Ahmed et al., 1995). O conteúdo é então comparado com os perfis de interesse dos usuários na comunidade, seguindo novamente o algoritmo de proximidade entre vetores de termos. Se for identificada uma alta proximidade entre os termos do conteúdo ou do documento submetido e o perfil de algum usuário, o OntoShare gera uma notificação informando onde este conteúdo pode ser obtido, quem o está compartilhando e quais foram os meta-dados adicionados. O perfil do usuário que está compartilhando o documento também é comparado com os conceitos relevantes do conteúdo que não estejam cadastrados, para estes últimos possam ser sugeridos e adicionados ao seu perfil. Essa característica é apresentada como uma capacidade de aprendizado do OntoShare, que evolui adaptativamente seu conhecimento sobre os usuários observando sua interação com o sistema (Davies et al., 2003).

Para a descrição do conteúdo uma outra ontologia é usada. Esta possui informações sobre palavras-chave associadas, o resumo do documento, seu título, a anotação fornecida pelo usuário, sua URL, o responsável por sua submissão e a data de armazenamento.