

3

Merging de Ontologias para uma Federação de Ontologias

No capítulo anterior descrevemos uma abordagem para facilitar a interoperabilidade entre sistemas de informação - a federação de bancos de dados.

Segundo (Brauner, 2005), de maneira geral, para interoperar, os sistemas de informação devem ser capazes de atender aos seguintes requisitos:

- Localizar as fontes de dados,
- Acessar as fontes de dados,
- Interpretar e processar os dados contidos nas fontes, sendo a heterogeneidade das fontes um dos maiores desafios enfrentados para conseguir realizar este requisito.

Este capítulo introduz a idéia de uma federação de ontologias como alternativa para promover a interoperabilidade entre fontes de informação. A abordagem de federação de ontologias aqui apresentada terá por base a abordagem de federação de bancos de dados.

A idéia é partir do conhecimento já explorado pela comunidade de banco de dados para integração de dados distribuídos e heterogêneos e traçar um paralelo para integração semântica destes mesmos dados através da integração das ontologias que representam o domínio onde estes dados estão inseridos. No final do capítulo será apresentada uma aproximação para uma federação de ontologias.

3.1 Federação de Ontologias

Como já dissemos, a Web é constituída por uma variedade de bases de informação heterogêneas e distribuídas formando uma grande rede de informações. Esta grande quantidade de informações disponíveis junto com novas tecnologias de comunicação, avanços em armazenamento digital e computação tem proporcionado à comunidade científica, tomadores de decisão e ao público de uma maneira geral, oportunidades antes nunca vistas de aquisição de conhecimento. Uma transformação radical tem acontecido com algumas áreas científicas como, por exemplo, as ciências biológicas, que tem acompanhado o

surgimento de novas áreas de pesquisa como a Bioinformática e a Biologia Computacional. O uso efetivo desta quantidade crescente de dados provenientes de bases de informação descentralizadas determina mudanças significativas na prática (Castilho et al., 2003).

Tendo em vista que os repositórios de dados são grandes em tamanho, dinâmicos e distribuídos fisicamente não é desejável nem factível garantir que todos os dados estejam em uma grande e única base centralizada para análise, pois serão necessários vários algoritmos a fim de extrair de maneira eficiente as informações relevantes das bases distribuídas, algoritmos estes nem sempre triviais.

Outro fato a ser considerado é que as bases de dados são criadas e operadas de maneira autônoma. Desta forma a quantidade de operações que podem ser realizadas nas bases de dados como, por exemplo, tipo de consultas permitidas e o modo preciso de garantir interação podem ser um pouco diversas para cada base demandando estratégias específicas para se obter a informação desejada dentro de regras operacionais impostas pela base de dados. Um exemplo de informação que se encaixa neste caso pode ser o de gerar estatísticas para realizar a mineração de dados.

Em muitas aplicações, como por exemplo, aplicações voltadas para a descoberta científica, os usuários precisam analisar os dados em diferentes contextos sob diferentes perspectivas. Não existe, nesta situação, uma ontologia universal que possa servir às necessidades de todos os usuários, ou ainda a um mesmo usuário em todo o contexto a ser considerado. Conseqüentemente, são necessários métodos para extração dinâmica de informação dependente de contexto e, integração de informação a partir de bases de dados distribuídas e heterogêneas representadas em ontologias especificadas pelo usuário para suporte à aquisição do conhecimento e tomada de decisão. As bases podem ser heterogêneas. Estas heterogeneidades podem ser de estrutura, de SGBDs, sintática e semântica. Cada base, implicitamente ou explicitamente, pode usar sua própria ontologia para representação dos dados.

Com o uso crescente de ontologias, o problema de sobreposição do conhecimento em um domínio comum ocorre muito freqüentemente. Ontologias de domínio específico estão sendo modeladas por múltiplos autores com pontos de vista diferenciados para representar os mesmos ambientes.

A comunidade de Banco de Dados tem sugerido um grande número de propostas com o objetivo de suportar comunicação global e compartilhamento do conhecimento. Entre estas propostas está o uso de Federação de Bancos de Dados para promover a integração. Em uma das aproximações para integração de dados via federação de dados, a resposta a uma consulta realizada pelo usuário é reunida após uma consulta realizada diretamente nas bases de origem. Os resultados da consulta acompanham a atualização das bases de origem, uma vez que a consulta realizada pelo usuário é diretamente direcionada a estas bases. Esta é a aproximação centrada em consulta assim chamada por (Hapimpour, 2004). Na outra aproximação de integração de dados por meio de uma federação de dados onde é gerado um esquema global, e que se aplica em casos em que os esquemas não sofrem mudanças constantes, os diversos bancos de dados formam a ontologia de uma comunidade específica pelo mapeamento das diversas bases de dados. Esta aproximação é a que mais se aplica ao nosso problema.

No nosso caso os bancos de dados distribuídos não formarão uma ontologia da comunidade. Cada banco de dados independente está representado por sua própria ontologia. Desta forma, temos várias ontologias que precisam ser integradas a fim de formar a ontologia global que, como em Banco de Dados, será fruto do mapeamento das diversas ontologias.

O mapeamento das diversas ontologias para a geração da ontologia global, será obtido através do *merging* das ontologias que fundirá as diferentes ontologias sobre o mesmo assunto em uma ontologia simples, que será a unificação de todas elas. Observamos que esta decisão, a de integrar as ontologias usando o *merging*, reside no fato de que com o uso de ontologias podemos usar uma aproximação lógica que tem sido usada com sucesso na direção de representação do conhecimento. A forma para integração de ontologias é muito geral e podemos levar em conta somente a estrutura das ontologias envolvidas e uma teoria em *Description Logic* (Baader et al., 2003). Como já foi dito, vamos realizar a integração semântica de dados com uma metodologia logicamente fundamentada.

3.2 Aproximações para uma Federação de Ontologias

Em um sistema de integração de dados uma consulta realizada pelo usuário precisa ser decomposta num conjunto de operações que corresponderão à informação que precisa ser reunida de cada base de origem e em seguida transformada para ser retornada ao sistema. Para que isto possa ocorrer este sistema de integração de dados precisa suportar dois tipos de operações básicas: *Get*() para consultar as bases de informação e *Transform*() para mapear os resultados na forma desejada. Estas operações precisam ser capazes de trabalhar com diferenças sintáticas e semânticas entre o vocabulário do usuário que busca por informação (usuário da ontologia global) e o vocabulário entendido por cada base (ontologias locais).

Segundo (Castillo et al, 2003), existem duas aproximações básicas para se trabalhar com diferenças semânticas entre uma ontologia global e ontologias locais. Uma delas é chamada de aproximação centrada nas bases de origem e a outra é chamada de aproximação centrada na consulta. No caso de aproximação centrada nas bases de origem cada base individual determina como os conceitos em uma ontologia local (específica da base) serão mapeados para os conceitos em uma ontologia global. Desta forma o usuário tem um controle reduzido do significado dos conceitos definidos na ontologia global e conseqüentemente no resultado da consulta. Ou seja, a semântica está centrada nas bases. Isto livra o usuário ou administrador do sistema de integração, da tarefa de especificar as transformações entre os conceitos locais e os conceitos globais: estas transformações são especificadas pelas bases locais.

De maneira oposta acontece na aproximação de integração de informação centrada na consulta. Os conceitos da ontologia global são definidos com base nos conceitos das ontologias locais (ontologia específica de cada base). Esta aproximação é mais adequada para aplicações de integração de dados nos quais os usuários precisam ter a habilidade para impor as ontologias (e conseqüentemente a semântica), de sua escolha para analisar e interpretar de maneira flexível as informações provenientes das bases autônomas. Isto requer um usuário ou administrador do sistema de integração para especificar de maneira precisa quais conceitos globais podem ser compostos a partir dos conceitos locais.

Apresentadas as aproximações acima para integração de sistemas de informação ressaltamos que não é nosso objetivo aqui discutir as vantagens e desvantagens de cada uma e em qual situação cada uma mais propriamente se aplica. Apenas foram aqui citadas para contextualização da nossa proposta de integração que parte de ontologias distribuídas representando bases heterogêneas que serão integradas gerando uma ontologia global, ou seja, de acordo com as propostas descritas nos parágrafos anteriores adotaremos neste trabalho a aproximação centrada nas bases para realizar a integração das ontologias na construção da nossa federação de ontologias.

3.3 *Merging* de Ontologias para definir um Sistema Federado Autônomo

O ponto chave para criar uma Federação de Ontologias é o processo de integração ou *merging* de ontologias específicas. O processo de *merging* de ontologias tem como *input* duas ou mais ontologias base e retorna uma ontologia integrada.

Como dissemos anteriormente o processo de *merging* das ontologias adotará uma aproximação centrada nas bases de dados, ou seja, será uma aproximação *bottom-up* que tem como *input* ontologias que serão integradas.

O processo manual de *merging* usando ferramentas de edição de ontologias é de difícil execução e sujeito a erros. Vários sistemas e *frameworks* para *merging* de ontologias têm sido propostos no suporte à engenharia do conhecimento como em (Hovy, 1998) e (Chalupsky, 2000). Estas aproximações lidam com heurísticas de mapeamento sintáticas e semânticas, que são derivadas a partir da crença de engenheiros de ontologias quando confrontados com a tarefa de conseguir o *merging* das ontologias, ou seja, comportamento humano é simulado. Estas aproximações não oferecem uma descrição estrutural do processo de *merging* global, apesar de usarem diferentes tipos de lógicas para comparação. Já (Stumme & Maedche, 2001) propõem um novo método chamado FCA-Merge para realizar o *merging* de ontologias. Eles seguem uma aproximação *bottom-up* que oferece uma descrição global do processo de *merging*. Para as ontologias base são extraídas instâncias a partir de um dado conjunto de documentos texto de um domínio específico por aplicação de técnicas de linguagem natural. Em seguida

são aplicadas técnicas matematicamente fundamentadas tiradas da Análise Conceitual Formal (Ganter et al, 2000) para derivar uma cadeia de conceitos como um resultado modelado do FCA-Merge. O resultado produzido é explorado e transformado para a ontologia resultante do *merging* por um engenheiro de ontologias. A extração de instâncias a partir de documentos-texto resolve o problema de que em muitas aplicações não existem objetos que são simultaneamente instâncias das ontologias bases e que poderiam ser usados como base para identificação de conceitos similares.

Não é nosso propósito aqui o desenvolvimento de nenhuma ferramenta para realizar o *merging* das ontologias que serão usadas no nosso estudo de caso. É sim nosso propósito, a integração semântica de dados por meio do *merging*. Estamos interessados no produto final dessa integração. Para realizar o *merging* usaremos a ferramenta *iPrompt* de (Noy & Musen, 2000). Uma ontologia global será gerada como resultado da integração das ontologias componentes de maneira similar ao que é feito na integração de bancos de dados utilizando a tecnologia de federação de bancos de dados.

3.4 Uma Arquitetura para Federação de Ontologias

De maneira similar ao que é feito em federação de bancos de dados, vamos definir uma arquitetura para federação de ontologias. A arquitetura que definiremos terá por base a aproximação de (Sheth & Larson, 1990) para uma federação de bancos de dados.

É uma arquitetura de 5 camadas (Figura 2). Ela é uma extensão da arquitetura de três camadas ANSI/SPARC com duas camadas adicionais. A arquitetura adotada consiste principalmente de:

1. Ontologias Locais (modelos conceituais de sistemas autônomos), cada um deles com sua específica ontologia base ou repositório de metadados;
2. Ontologias normalizadas (transformação das Ontologias locais em um modelo de dados comum);
3. Ontologias de exportação (visões da ontologia normalizada que descrevem as partes relevantes da ontologia para a federação);
4. Uma Ontologia integrada (Ontologia Global derivada a partir da combinação dos esquemas de exportação) e;

5. Aplicações diferentes em uma camada acima (camada do esquema externo, que usa uma ontologia externa com suas visões específicas).

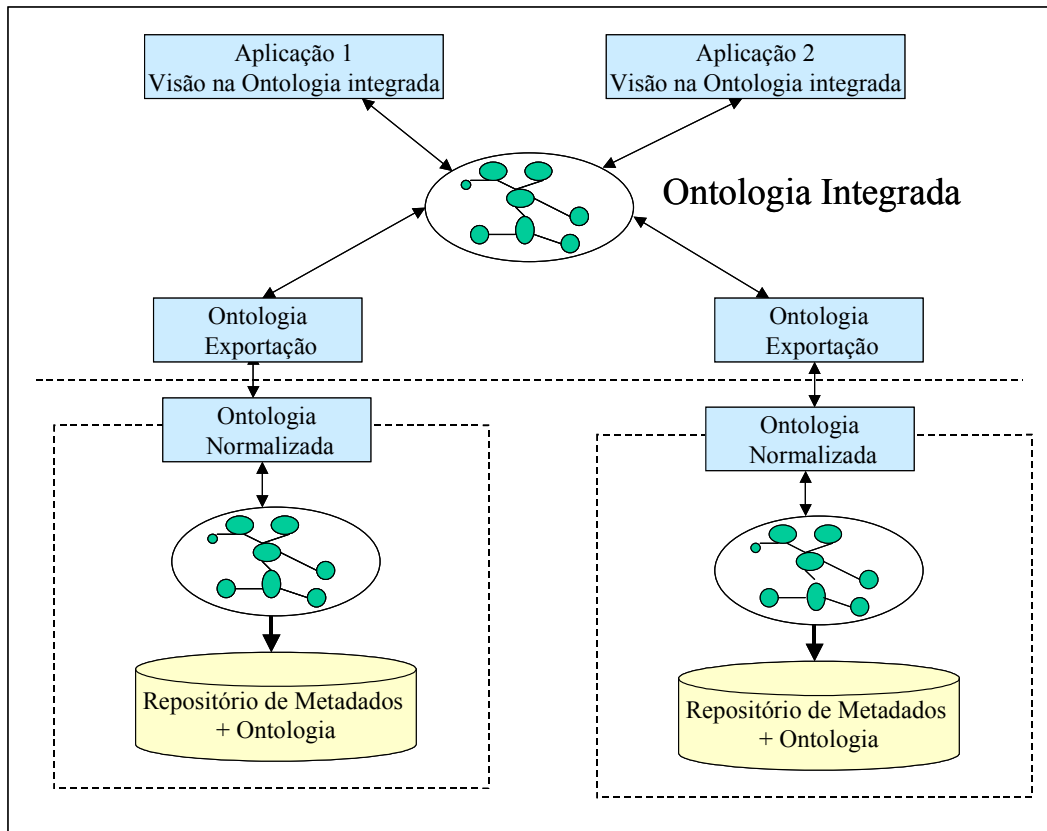


Figura 2 : Arquitetura para uma federação de ontologias.