

5

Conclusão e trabalhos futuros

Neste capítulo fazemos uma retrospectiva do trabalho realizado, uma avaliação da proposta de solução de integração de dados ou conhecimentos mostrada na dissertação e também fazemos uma breve comparação desta abordagem com outras soluções sugeridas na literatura. No final as principais contribuições e indicações de futuros trabalhos são também apresentadas.

5.1 Retrospectiva

Neste trabalho apresentamos a integração semântica de dados através do *merging* de ontologias fazendo uma analogia à integração de esquemas de dados heterogêneos e distribuídos em um modelo global de uma federação de banco de dados.

Inicialmente apresentamos um histórico do problema de integração de dados que tem sido até hoje a questão básica de pesquisa na área de banco de dados. Dentre esses problemas centralizamos nosso foco nas técnicas de integração em federações de bancos de dados já que são as mais similares ao *merging* de ontologias que é o foco principal deste trabalho.

Em seguida apresentamos o problema de integração de dados distribuídos na Web e, a solução encontrada para a integração destes esquemas de dados representados agora em ontologias. Esta solução, o *merging* de ontologias para a criação de uma federação de ontologias, foi então detalhada juntamente com uma arquitetura para a federação de ontologias.

Visando compreender mais profundamente a solução de integração por *merging* de ontologias apresentamos um estudo de caso onde detalhamos o processo e verificamos algumas vantagens desta abordagem.

Finalmente faremos a seguir uma análise da solução adotada, uma comparação com outras técnicas e um resumo de contribuições desta dissertação bem como sugestões para continuações do trabalho.

5.2 Avaliação da estratégia usada

Na arquitetura de integração de dados usando federação de dados, os diferentes esquemas que são integrados para formar o modelo global possuem diferentes mecanismos de segurança. As dificuldades resultantes do controle de autorização neste tipo de ambiente residem no fato de que os objetos e os usuários são distribuídos.

Nos sistemas relacionais (contexto do nosso estudo de caso) as autorizações podem ser controladas de modo uniforme por administradores de bancos de dados que podem especificar construções de alto nível fazendo uso do SGBD, uma vez que funções para este fim, já estão originalmente implementadas em muitos dos SGBDs relacionais existentes no mercado hoje (podemos citar como exemplo DB2 e ORACLE). A mesma facilidade não acontece no caso geral, onde o esquema Global não é baseado no modelo relacional e o modelo adotado para o esquema Global pode não ser estruturado (como por exemplo, arquivos em XML). Nestes casos o uso de Ontologias para especificação dos mecanismos de segurança pode ajudar a minimizar o problema. De posse de uma Ontologia que modele os mecanismos de segurança, pode-se utilizar um provador de teoremas para se verificar certos aspectos de segurança, ou seja, o uso de uma linguagem lógica (como OWL-DL) para descrever o modelo global permite que se construa uma ontologia de *controle de autorização* na própria linguagem. Problemas como autenticação remota de usuários, gerência de regras de autorização distribuídas bem como o tratamento de visões e de grupos de usuários (usuários de um grupo podem estar localizados em vários sites e o acesso a um objeto pode ser concedido a vários grupos, que por sua vez são distribuídos) podem já estar definidos na própria ontologia que representa o modelo global. Com o uso de uma linguagem lógica para representar o “modelo de consulta da federação” pode-se provar propriedades importantes como a segurança do controle de acesso, ou seja, pode-se “provar” que certo tipo de usuário não tem autorização para acesso a um conceito da federação. Portanto podemos concluir que esta solução de uso de uma federação de ontologias para promover a integração de dados, no caso de controle de acesso é mais efetiva.

Na arquitetura de federação de dados adotada para promover a interoperabilidade, podem ocorrer diversos tipos de heterogeneidades, estes tipos

de heterogeneidades foram citados no capítulo 2 - heterogeneidade quanto a SGBDs, estrutural, sintática e semântica. Cada base de dados local provê uma descrição dos dados a serem exportados (esquemas locais) para se promover um esquema global onde os usuários estarão providos com uma visão única. (Hass, 2003), destaca a heterogeneidade semântica como um dos problemas que podem dificultar a integração de dados usando federação. No mapeamento dos esquemas locais para a geração do esquema de exportação é muito difícil identificar associações, onde termos iguais podem ter diferentes significados, ou termos diferentes podem ter os mesmos significados ou ainda se ter estruturas diferentes representando interpretações diferentes.

Na arquitetura por nós sugerida, integração de dados através de uma federação de ontologias, o modelo de dados global será uma ontologia resultante do *merging* de ontologias. Esta ontologia está especificada em uma linguagem lógica (OWL-DL). Desta forma a integração dos dados pode ser realizada usando-se apenas a representação dos termos no nível lógico, ou seja, independente da implementação dos mesmos.

Por exemplo, se tivermos um conceito não primitivo que não estará representado no modelo físico do banco de dados (não implementado), no uso de federação de dados para a integração, este conceito só estará acessível via consulta, ou seja, será necessário fazer uso de uma linguagem de consulta para que a informação seja obtida isto demanda tempo de acesso às bases de dados de origem para a obtenção da resposta. A complexidade pode aumentar caso seja preciso fazer uso de algoritmos sofisticados de acesso.

A integração das bases representadas por ontologias vai nos dizer como um conceito global se remete a um ou mais conceitos locais, ou seja, o modelo global pode ser validado sem a necessidade de implementação. Isto facilita muito quando da necessidade de manutenção do modelo, pelo fato deste ser representado em uma linguagem lógica (OWL-DL) fica fácil a manutenção ou reutilização de parte do modelo.

5.2.1 Vantagens do uso de ontologias

O uso de Ontologias torna possível definir uma infra-estrutura para integrar sistemas inteligentes no nível lógico. O nível lógico é independente do nível de implementação. Ontologias apresentam grandes vantagens como:

- Colaboração: possibilitam o compartilhamento do conhecimento entre os membros interdisciplinares de uma equipe.
- Interoperabilidade: facilitam a integração da informação, especialmente em aplicações distribuídas.
- Informação: podem ser usadas como fonte de consulta e de referência do domínio;
- Modelagem: as ontologias são representadas por blocos estruturados que podem ser reusáveis na modelagem de sistemas no nível lógico.

5.3 Contribuições

Neste trabalho mostramos como, com o uso de uma ferramenta de construção e manipulação de ontologias (Protégé), juntamente com um raciocinador para *Description Logic*, pode-se implementar um esquema de integração federado de ontologias no estilo do OBSERVER (Mena, 1996), (Silvestre, 2005). O mapeamento de consultas do modelo global para os esquemas locais via esquema de exportação foram no nosso caso compilados no esquema global - ontologia global mais especificamente, em vez de serem realizados a cada consulta. Até mesmo a camada de aplicação (como modelo de autorização de nosso estudo de caso) é compilada em conjunto na ontologia global.

Quando se tratar de federação de bases de dados, com distribuição das mesmas, as consultas na ontologia podem ser mapeadas em consultas às respectivas bases, usando o esquema de exportação como guia de geração das mesmas. Entretanto, este processo não foi estendido para base de dados como mencionado acima, por estar fora do escopo da presente dissertação. O mapeamento e geração de consultas é assunto tratado pela comunidade de banco de dados já há algum tempo.

Mostramos ainda como de posse de uma federação de ontologias, certos conceitos locais podem ser aproveitados na construção de uma camada de

aplicação, podendo esta, ainda ser compilada na própria linguagem descritora do esquema global (OWL-DL).

Concluimos que o uso de Ontologias e *Description Logic* na integração de modelos conceituais heterogêneos pode ser proveitoso, principalmente através do uso de um raciocinador (como o *Pellet* no nosso estudo de caso) para verificar a consistência do modelo global e inferir propriedades interessantes na camada de aplicação.

Finalmente, o uso de ontologias mostrou-se mais interessante de ser aplicado na integração de base de dados (ontologias), mesmo que seus respectivos modelos conceituais tenham que ser transportados para as respectivas ontologias locais. O ganho em flexibilidade e no uso de uma única linguagem para a construção da federação, assim como na construção de eventuais camadas de aplicação, demonstrado no nosso estudo de caso, parece corroborar esta conclusão.

5.3.1 Aplicação da solução

Esta solução de federação de ontologias poderia ser aplicada no controle de acesso em um data warehouse definido através de uma federação bancos de dados. (Samos et al., 2000) define uma arquitetura para data warehouse baseada em federação de dados. Ele admite ser o processo de autorização de acesso em um data warehouse bastante crítico, uma vez que os dados armazenados em um sistema de integração deste tipo são utilizados para suporte à tomada de decisões, ou seja, tem a confiabilidade como uma característica fundamental. Estes dados precisam estar fortemente protegidos contra acesso não autorizado. Ele sugere a adoção na sua arquitetura de federação de data warehouse, de um *esquema de autorização* para proteção aos dados.

Este esquema de autorização deverá ser usado para dar apoio ao controle de inferência para se obter uma informação. O problema do controle de inferência aparece porque um usuário pode obter informação não autorizada, ou ainda conhecimento sobre a existência de uma informação não autorizada, através do acesso a dados autorizados, ou seja, quando existe um fluxo de informações a partir um objeto classificado no nível N_1 para outro objeto classificado no nível N_2 e $N_2 < N_1$. Este problema é piorado se levamos em conta as abstrações semânticas do modelo de dados.

Podemos resolver este problema aplicando a arquitetura sugerida neste trabalho. Os diversos Data Marts representados nas suas ontologias locais seriam integrados para formação do modelo global do data warehouse que poderia ter uma hierarquia de classes definindo o controle de acesso às informações de cada base, hierarquia esta definida em uma linguagem lógica (OWL-DL).

Outra situação em que a nossa solução se encaixa, foi vivenciada em nossa experiência de mercado. Um grande problema enfrentado hoje pelas grandes corporações é da definição única do que seria o seu cliente corporativo. Num contexto de uma empresa que comercialize seguro, por exemplo, um cliente que faça parte de uma base de seguro de saúde pode fazer parte de outra base independente de seguro de veículo e ainda ser um cliente potencial para adquirir um outro tipo de seguro que a empresa comercializa. Seria necessário então traçar um perfil mais completo do cliente para que ele pudesse ser mais bem atendido e para que a empresa pudesse também ter ciência de que tipo de produtos ele tem necessidade e que ainda não possui. Para tanto seria necessária à integração destas bases de dados. O uso de ontologias para promover a integração seja com o uso de uma linguagem lógica permitiria inferência para colher dados sobre cliente e poder classificá-los, ou seja, esta é uma contribuição substancial e bastante aplicável em um problema real.

O nosso estudo de caso se deu em uma pequena amostra considerando em princípio duas pequenas bases de dados. A nossa intenção era a ilustração da viabilidade da solução. Na análise do resultado ficou evidente para nós de que esta solução é perfeitamente aplicável em um caso real onde teríamos um número maior de variáveis a serem analisadas. Aplicar esta solução em um ambiente real implica em um maior esforço na definição das ontologias a serem integradas. O processo para geração do modelo global e da camada de aplicação seria o mesmo implicando somente na definição de um maior número de regras lógicas usando OWL-DL.

5.4 Trabalhos futuros

- Geração de consultas SQL a partir de deduções em Description Logic. Uso na implementação da federação de ontologias quando um

esquema global “faz menção” a um conceito definido (não primitivo).

- Utilização desta arquitetura para geração do modelo do data warehouse a partir dos data marts distribuídos.

Utilização de ontologias para ampliar conhecimento sobre dados a serem integrados para realização de mineração de dados.