

3

A Abordagem Adotada e Trabalhos Relacionados

A arquitetura proposta resulta de uma evolução de estudos e trabalhos anteriores em que outras abordagens para integração de dados foram discutidas no contexto de *e-learning*, considerando especialmente o caso do projeto PGL. Neste capítulo apresentamos um resumo destes estudos e trabalhos anteriores, descrevendo as alternativas de integração e discutindo sua aplicabilidade no contexto desejado. Discutimos também alguns trabalhos relacionados que ressaltam características importantes para a implementação de nossa abordagem.

3.1. Alternativas para Integração de LOs

Conforme descrito anteriormente (seção 1.3), este trabalho tem como objetivo propor uma arquitetura para compartilhar LOs a partir de seus metadados em um ambiente distribuído e heterogêneo. Para prover este compartilhamento é necessário interconectar os repositórios de metadados dos diferentes sítios que compõem a comunidade de *e-learning*, ou seja, as instituições que firmam uma parceria com o intuito de promover o compartilhamento de LOs. Ao longo desta seção, são discutidas as principais abordagens para integração de dados, feitas considerações e apresentados trabalhos relevantes no contexto de *e-learning* que a implementam.

3.1.1. Centralizada

A centralização pode ser observada através dos repositórios globais de materiais instrucionais, que são cada vez mais comuns na Internet, tais como o TeleCampus [55], o *Campus Alberta Repository of Educational Objects* (CAREO) [56], *Repository of Objects with Semantic Access for e-learning* (ROSA) [57] e o *Multimedia Educational Resource for Learning and Online Teaching* [58]. Os repositórios globais, geralmente, mantêm *links* para conteúdos

de aprendizagem que se encontram na *Web*, embora alguns deles armazenem fisicamente os conteúdos.

Outra abordagem interessante de centralização é o *EducaNext/Universal Project* [59]. EducaNext é um portal multilíngue para compartilhamento, recuperação e reuso de recursos didáticos. É um resultado do UNIVERSAL Project, que provê serviços globais para troca de recursos didáticos entre instituições de ensino superior, usando a *UNIVERSAL Brokerage Platform*. Deste modo, há uma centralização dos serviços de catalogação e disponibilização dos recursos didáticos.

No início deste trabalho foi estudada uma abordagem centralizada por facilitar a busca e gerência dos LOs, provendo um controle melhor, com rápida indexação e recuperação dos objetos. Desta forma, uma cópia dos metadados seria extraída dos sistemas locais e armazenada num banco de dados centralizador, numa idéia similar a *data warehouses* [60], embora tendo num escopo diferente.

Apesar da facilidade oferecida, a grande desvantagem em manter um banco de dados centralizado para sistemas de *e-learning* é a dificuldade na coordenação desta centralização e uma possível perda de autonomia dos sítios participantes. A questão da autonomia dos sítios é importante, pois, em geral, as instituições não estão interessadas em perder sua independência e controle.

3.1.2. Distribuída com Replicação

De forma a preservar a autonomia local e promover a colaboração entre os sítios, foi estudada a abordagem distribuída com replicação. A implementação desta abordagem no cenário de comunidades de *e-learning* levaria as instituições participantes a manterem seus recursos didáticos (.doc, .ppt, .pdf, etc.) localmente, todavia, os metadados dos mesmos seriam replicados mutuamente.

Isto seria uma alternativa interessante uma vez que o volume de informação dos metadados é consideravelmente menor do que o utilizado para representar e armazenar os próprios recursos didáticos. Deste modo, haveria em cada sítio uma visão global dos metadados referentes a todos LOs compartilhados na comunidade. Como dito anteriormente, os metadados constituem o ponto de

acesso para os recursos didáticos e, sendo assim, a idéia seria obter o ponteiro para qualquer LO de qualquer fonte através de uma consulta local.

Esta alternativa chegou a ser desenvolvida como parte inicial desta dissertação, através da definição de uma arquitetura [61], a qual tratava a replicação dos metadados nos sítios de uma comunidade de *e-learning*. Apesar de oferecer uma busca local, a necessidade do controle de replicação, bem como o armazenamento adicional de metadados gerou resistência entre alguns membros participantes da comunidade do estudo de caso (descrito na seção 6.1). Tal fato fez com que esta alternativa não fosse adotada neste trabalho.

3.1.3. Par-a-Par

Outra abordagem estudada foi a Par-a-Par (*Peer-to-Peer* – P2P), onde os membros de uma comunidade seriam tanto provedores como demandadores de recursos (serviços e conteúdo).

Um exemplo de aplicação bem sucedida do modelo P2P em ambientes de *e-learning* está presente no projeto Edutella [62], o qual propõe uma infra-estrutura P2P baseada em RDF para conectar pares educacionais heterogêneos com diferentes tipos de repositórios, linguagem de consulta e diferentes tipos de esquema de metadados. Cada par do Edutella é capaz de desenvolver um número básico de serviços tais como consulta, replicação e mapeamento.

As comunidades de *e-learning* são formadas tradicionalmente por instituições que detêm o seu conteúdo didático armazenado localmente e mantêm aplicações legadas fazendo acesso a este conteúdo. A estrutura presente nessas instituições corresponde, em sua maioria, à arquitetura cliente-servidor. Isto significa que para projetar uma arquitetura P2P é necessário prover uma reconfiguração do ambiente, o que constitui um fator de dificuldade à sua implantação. Além disso, a disponibilização dos recursos de forma a debilitar o desempenho das aplicações locais, bem como questões de segurança dos dados e da rede também são fatores contrários à adoção desta alternativa. Sendo assim, a implantação do modelo P2P como alternativa para integração de objetos de aprendizagem da comunidade do estudo de caso foi descartada.

3.1.4. Distribuída Baseada em Mediadores e Tradutores

Esta abordagem visa preservar a autonomia dos sistemas locais, possibilitando o compartilhamento e a troca de informações entre fontes de dados autônomas e heterogêneas.

A arquitetura de mediadores efetua o acesso aos dados distribuídos em múltiplas fontes de informação através de consultas que são submetidas ao sistema, via mediador, e este as transforma em sub-consultas a serem enviadas às fontes. As sub-consultas geradas pelo mediador devem ser traduzidas para linguagens de consulta de cada fonte, através de tradutores. Ao final, os resultados das consultas são traduzidos para a linguagem de consulta do mediador e a resposta é devolvida ao usuário.

A arquitetura baseada em mediadores e tradutores tem sido amplamente aceita para resolver problemas de integração de informação, especialmente para aplicações baseadas na *Web*. Conforme discutido no capítulo anterior, as características de SIBM são necessárias às comunidades de *e-learning*, pois a ênfase na integração de LOs está na busca e recuperação, sendo que as atualizações devem ocorrer localmente; o uso de um esquema federado com um modelo integrador dos padrões de metadados (e de outros possíveis meta-esquemas) é desejável; o ambiente de integração deve permitir o acesso a componentes estruturados, semi-estruturados e não estruturados; além da própria heterogeneidade inerente a estas comunidades.

A abordagem usada pela arquitetura proposta nesta dissertação faz uso da alternativa de integração baseada em mediação com o suporte da tecnologia de serviços *Web* (*web services*) [63] e ontologias [64]. O uso de serviços *Web* provê uma forma padronizada e interoperável de integrar sistemas existentes ou aplicações, tornando-se mais adequada à plataforma *Web*. As ontologias provêm a interoperabilidade semântica entre as instituições.

Na seção seguinte, são apresentados alguns trabalhos que tratam desta abordagem, considerando as tecnologias de serviços *Web* e ontologias. O Anexo B apresenta outros trabalhos de integração de dados, enfatizando aqueles que são baseados na arquitetura de mediadores e tradutores.

3.2. Trabalhos de Integração

Uma vez definida a abordagem a ser adotada, passamos a discutir trabalhos relacionados que acrescentam características importantes à solução proposta escolhida e implementada, tais como o uso de serviços *Web* para promover a interoperabilidade e flexibilidade do ambiente de integração, o uso de ontologias para promover um tratamento melhor da heterogeneidade semântica e a implementação de uma arquitetura de mediadores em multicamadas para lidar com os diferentes padrões/meta-esquemas de LOs.

3.2.1. Projeto Integra

O Integra [65] é um sistema, baseado em mediação, para integração de informações distribuídas em fontes de dados na *Web*. Neste sistema, são adotados conceitos, tais como metadados, contextos e ontologias na resolução da heterogeneidade da informação. Estes conceitos são utilizados para identificar correspondências e resolver conflitos semânticos entre informações de diversas fontes de dados heterogêneas na *Web*. O sistema de integração Integra possui uma arquitetura baseada em mediação que adota a abordagem *GaV* (discutida no capítulo anterior) na definição de mapeamentos entre o esquema de mediação e os esquemas das fontes de dados. É utilizada a linguagem XML como modelo comum para intercâmbio de dados e XML *Schema* como representação padrão para o esquema de mediação e esquema das fontes de dados.

O sistema Integra foi originalmente proposto para resolver aspectos sintáticos e estruturais na integração de informações, mas sua arquitetura foi estendida de forma a incluir aspectos para o tratamento semântico da informação.

No sistema Integra é utilizada uma ontologia de domínio para uma interpretação semântica dos conceitos encontrados em cada uma das suas fontes de dados. Os conceitos semanticamente similares são identificados nas fontes e posteriormente integrados. Uma das características do Integra é a forma de organização de metadados do sistema agrupados através de contextos. Um contexto contém metadados relacionados ao seu significado, propriedades (tais como fonte, qualidade e precisão) e organização [66][67]. Em um sistema de

integração de informações, contextos podem conter descrições sobre a natureza estrutural, organizacional e semântica das fontes de dados distribuídas.

Em nosso trabalho, temos o mesmo contexto (*e-learning* e LOs) em todos os sítios, mas uma vez que diversos padrões podem ser usados e cada padrão pode ser implementado em um esquema e modelo de dados diferentes, então continua sendo necessário lidar com a heterogeneidade.

3.2.2. Projeto Edutella

O projeto Edutella [68] provê metadados baseados em RDF em uma infraestrutura P2P para a troca de recursos educacionais usando padrões como IEEE LOM, IMS e ADL SCORM para descrever materiais de curso. O projeto Edutella implementa consultas distribuídas a partir de um modelo de dados comum (chamado de ECDM) baseado em *Datalog* e linguagem de consulta RDF-QEL.

Para permitir um par participar na rede Edutella, faz-se necessária a utilização de tradutores baseados tanto no modelo de dados quanto na linguagem de consulta comum. Para a comunicação na rede Edutella, o tradutor transforma o modelo de dados local para o modelo de dados comum do Edutella e vice-versa. A conexão com a rede é feita através da arquitetura JXTA P2P, transformando as consultas baseadas no ECDM no formato RDF/XML.

O projeto Edutella consiste de serviços de consulta, serviço de anotação, serviço de registro e serviço de mediação. O registro de um par é baseado na informação do esquema, informando à rede que tipo de esquema o par utiliza e algumas possíveis restrições. A mediação ocorre utilizando um par especializado que funciona como mediador na rede Edutella distribuindo a consulta para os pares apropriados. O Edutella apresenta, entre outros componentes, um tradutor para o repositório (*Open Learning Repository*) que utiliza um subconjunto dos padrões de metadados IMS/LOM.

Apesar de não adotarmos a abordagem P2P, é importante observar este projeto devido às características de sua mediação, incluindo questões ligadas ao tratamento dos padrões de metadados para LOs.

3.2.3. Projeto ONES

O projeto ONES (*One-Stop eLearning Portal*) [69] tem como objetivo desenvolver um sistema que apóie a busca por cursos de interesse do aprendiz em universidades virtuais. Devido ao fato de existir heterogeneidade entre as universidades virtuais, o sistema ONES provê mecanismos para esconder a distribuição dos portais de *e-learning* e a heterogeneidade semântica entre as informações. De forma a alcançar esses objetivos, o sistema disponibiliza de tecnologias e paradigmas tais como serviços *Web*, mediadores, RDFS e OWL

Considerando a arquitetura de integração de dados, o portal de agregação funciona como um mediador que suporta uma visão virtual que integra diversas fontes de dados que são representadas por portais de *e-learning*. Uma vez que os portais de *e-learning* foram criados de forma independente, apresentam interfaces heterogêneas. Desta forma, o sistema ONES define tradutores que provêm uma interface para requisição de metadados dos LOs. Cada tradutor na arquitetura ONES funciona como um serviço *Web* que pode ser publicado e invocado por outros portais.

No sistema ONES foram desenvolvidos dois tipos de ontologias: Ontologia Representacional (que provê entidades representacionais sem estar relacionada a um domínio em particular) e Ontologia de Domínio (que captura um conhecimento válido para um tipo particular de domínio).

A ontologia de domínio foi dividida em duas sub-classes: Ontologia de meta domínio, que captura o modelo ou esquema do domínio; e a Ontologia de instância de domínio, que representa uma instância de uma ontologia de meta domínio.

No sistema ONES foram utilizadas as linguagens OWL, RDF Schema e ODL (*Object Definition Language*) como uma ontologia representacional. Além disso, foi definido uma ontologia de meta domínio chamada de *ontologia de objeto de e-learning* e uma ontologia de instância de domínio chamada *ontologia de assunto*, que descreve os termos que podem ser usados para os LOs.

A implementação da interação dos usuários com o sistema é através de HTML, operações GET, POST e PUT. Assim como a forma de apresentação, os LOs seguem a abordagem orientada a objetos e o uso da linguagem OQL (*Object*

Query Language) para a especificação das consultas. A comunicação entre os tradutores e as interfaces ONES consiste de mensagens SOAP, onde o mediador envia as consultas através de SOAP para as interfaces ONES e estas enviam as respostas para o mediador, que primeiro combina os resultados e então os entrega para o usuário.

Embora o enfoque deste projeto seja a busca por cursos, diferentemente do nosso trabalho que visa o compartilhamento de LOs, é interessante observar as soluções apresentadas em ONES, visto que também adotam serviços *Web* na implementação dos tradutores e trabalham com ontologias para o tratamento das heterogeneidades.

3.2.4. Projeto MIWEB

MIWeb (*Mediator-based Integration of Web Sources*) [70] é um sistema que contribui com uma arquitetura de mediador no contexto de fontes de dados na *Web* para busca de materiais de aprendizagem e publicações. O sistema MIWeb pesquisa em fontes de dados que descrevem tipos diferentes de documentos *Web*, tais como, a máquina de busca Google, a fonte científica Citeseer e recursos específicos para *e-learning* desenvolvidos no projeto *NewEconomy*. MIWeb utiliza padrões de metadados para integração e tecnologias como HTTP, RDF e a linguagem de consulta RDQL. A arquitetura MIWEB consiste de três componentes principais: mediador, tradutor e mapeador. No mediador, o modelo de dado utilizado é o modelo RDF e o esquema de metadados adere ao padrão LOM, utilizado para descrever os recursos de *e-learning*.

Os usuários consultam o sistema usando RDQL, uma linguagem de consulta para fontes em RDF. A consulta então é enviada ao componente mediador que divide a consulta em subconsultas e as envia para os componentes tradutores, que as transformam para uma representação específica da fonte em RDF.

O componente mapeador é usado pelos tradutores para resolver a heterogeneidade sintática e semântica entre o mediador e o tradutor, transformando dados RDF de um esquema específico da fonte para dados no modelo LOM. A transformação é baseada no mapeamento manual definido em XSLT.

O projeto MIWEB acrescenta um componente de mapeamento que é interessante também em nosso contexto. Entretanto, este projeto, difere de nossa abordagem por apresentar componentes menos flexíveis e um tratamento semântico menos formal, uma vez que não se baseia em serviços *Web* e ontologias.

3.2.5. Projeto IBHIS

IBHIS (*Integration Broker for Heterogeneous Information Sources*) [71] é um projeto com o propósito de prover um intermediário de integração (*broker*) permitindo o uso de um conjunto de fontes de dados heterogêneas e distribuídas. O projeto define uma arquitetura orientada a serviços, baseada no conceito de sistemas de banco de dados federados. O domínio da aplicação no IBHIS é aplicado no domínio social e da saúde.

A arquitetura do IBHIS é uma combinação de banco de dados federado e sistema baseado em serviços. IBHIS utiliza a tecnologia de serviços *Web* e padrões e protocolos abertos tais como, Java, SOAP, WSDL e UDDI. A visão global dos dados distribuídos é realizada através da criação de um ou mais esquemas federados de acordo com os requerimentos do usuário. Na arquitetura, o serviço de acesso aos dados (DAS) é construído usando serviços *Web*. O DAS é usado para prover acesso transparente às fontes de dados distribuídas, autônomas e heterogêneas. O DAS pode ser programado usando diferentes linguagens de programação e pode ser executado em diferentes sistemas operacionais, mas o DAS provê uma forma unificada para acessar os dados.

A arquitetura IBHIS utiliza um serviço de ontologias para recuperação de informações mantidas nas fontes de dados. O conhecimento é obtido através da inspeção dos esquemas das fontes. Em seguida, a ontologia é consultada para relatar sinônimos e conceitos semânticos e, a partir disto, os detalhes são passados para o serviço de integração de esquemas, o qual os transforma para um modelo comum. Os relacionamentos entre os componentes dos esquemas são identificados e é gerado um mapeamento apropriado. Este processo envolve resolver várias formas de heterogeneidade e resultar em um esquema federado. O

serviço de ontologia, neste caso, irá prover informações de domínio ajudando a resolver esta questão de heterogeneidade.

Apesar de ter um escopo diferente (domínio social e da saúde) e uma arquitetura de sistemas de banco de dados federados, este projeto é baseado em serviços *Web* e faz uso de ontologias para o tratamento de heterogeneidade. No próximo capítulo discutimos nossa proposta para a questão do tratamento da heterogeneidade.