

## 5 Arquitetura Proposta

Neste capítulo detalhamos a arquitetura proposta que provê acesso a fontes de dados autônomas, heterogêneas e distribuídas, as quais podem ser desde sistemas gerenciadores de bancos de dados até simples coleções de arquivos. É utilizada uma abordagem distribuída baseada em mediadores e tradutores, sendo implementada através de serviços.

### 5.1. Visão Geral da Arquitetura

A arquitetura utilizada no LORIS faz uso de um esquema federado ou esquema global cumprindo um papel um integrador passivo. Um integrador passivo [71] procura por informações em demanda e oferece uma visão customizada dos dados. Uma questão importante gerenciada pela arquitetura é o tratamento da heterogeneidade entre os padrões utilizados pelas fontes de dados (membros da comunidade) para representação dos metadados de seus recursos educacionais.

Os componentes da arquitetura, apresentados na Figura 5.1, são descritos a seguir:

- Fontes de Informação (SGBDR, BDOO, BDXML, Arquivos): são fontes de dados que podem ser autônomas e heterogêneas, visto que, na maioria das vezes elas não foram especificadas com a intenção de compartilhamento. As fontes de informação podem ser banco de dados, repositórios de objetos, bases de conhecimento, etc.
- Mediadores: são módulos de *software* que exploram o conhecimento representado num conjunto ou subconjunto de dados com o objetivo de gerar informações para aplicações residentes numa camada superior [45]. Podem ser vistos como uma camada intermediária entre a camada das aplicações e a camada das fontes de dados, tendo

como função a aplicação do conhecimento especializado a um domínio específico para agregar valor [45].

- Tradutores (Tr): também conhecidos como *wrappers*, convertem os dados das fontes distribuídas e heterogêneas segundo um modelo de dados comum e convertem as consultas da aplicação em consultas específicas da fonte de informação correspondente.

A arquitetura é baseada em serviços *Web* e, portanto, implementa um modelo que utiliza protocolos padrões abertos, tais como: Java, SOAP, WSDL e XML. A Figura 5.1 apresenta a arquitetura numa visão geral, sendo estruturada em três camadas: Camada de Aplicação, Camada de Mediação e Camada de Acesso aos Dados.

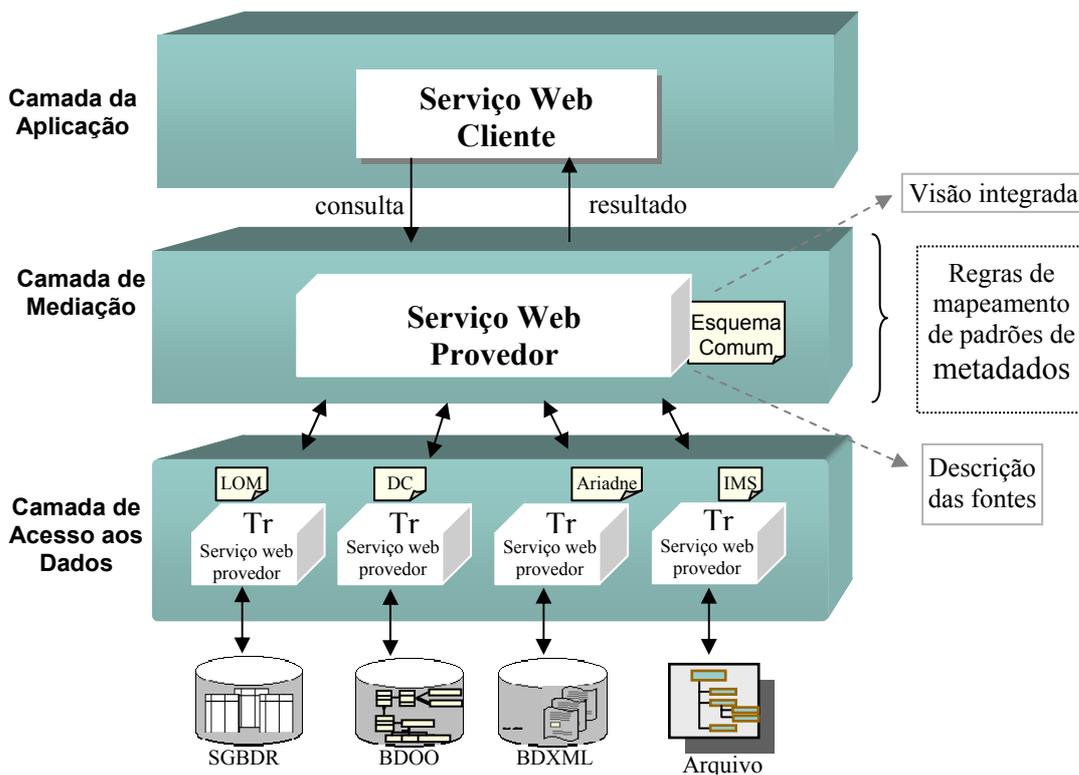


Figura 5.1: Uma Visão Geral da Arquitetura LORIS

## **5.2. As Camadas da Arquitetura**

As camadas da arquitetura são detalhadas nesta seção, onde são descritos os serviços relevantes.

### **5.2.1. Camada de Aplicação**

Esta camada permite às aplicações clientes obterem acesso à camada de mediação, a qual é encarregada de oferecer os serviços de consulta. A camada de aplicação provê duas escolhas para o usuário: uma interface de consulta comum e uma interface própria de uma aplicação externa. A interface própria é desenvolvida pela instituição participante da comunidade de *e-learning*, disparando o serviço disponibilizado pelo LORIS. Através das interfaces de consulta disponíveis, a instituição participante obtém acesso aos serviços de busca de LOs oferecidos pela camada de mediação.

Além da interface de consulta, também há uma interface para a administração dos participantes da comunidade. Através do uso desta interface pode-se invocar o serviço de registro de fontes, o qual é exemplificado no cenário em que uma nova instituição deseja fazer parte da comunidade. Sendo assim, o novo participante deve cadastrar-se através do serviço de registro, que será discutido em mais detalhes na seção 5.2.2.1.

### **5.2.2. Camada de Mediação**

Esta camada representa um papel intermediário entre a camada de aplicação e o acesso aos dados das fontes, podendo ser vista como uma provedora de serviços, que fornece serviços de mediação entre as fontes de dados que compõem a comunidade de *e-learning*. Nesta camada, uma visão global dos dados distribuídos é obtida através da criação de um esquema global ou esquema comum. O esquema global suporta uma visão virtual dos dados que integra as diversas fontes de dados (repositórios de LOs), mas não armazena esses dados integrados como é feito em *data warehouse*.

A camada de mediação é responsável por gerenciar todas as solicitações feitas pelo usuário e encaminhá-las para cada tradutor participante na arquitetura. É responsável também por disponibilizar uma API (*Application Program Interface*), que possibilita às diversas aplicações consultarem diferentes fontes de dados como se estivessem num ambiente local e homogêneo.

Para o funcionamento desta camada foram definidos:

- **Modelo de Dados**

O primeiro passo para a criação de um sistema de integração é prover uma homogeneização, isto é, definir o esquema da visão integrada através da integração dos elementos dos esquemas das fontes [82].

O esquema da visão integrada é a representação de um modelo global. No LORIS foi utilizada uma generalização dos conceitos envolvidos nos padrões de metadados para *e-learning*. O IEEE LOM [14] serviu de base para a abstração dos conceitos necessários e posterior construção do modelo global, uma vez que este padrão cobre grande parte dos metadados representados pela maioria dos padrões já existentes para descrição de recursos didáticos: Dublin Core, ARIADNE, IMS, etc.

Além de definir este esquema comum também foi necessário estabelecer correspondências entre o esquema global (comum) e os esquemas das fontes locais, ou seja, foram definidas regras de mapeamento de modo a garantir a visão integrada dos dados das fontes. Sendo assim, essas regras de mapeamento, representadas através de ontologias, permitem que o usuário de cada fonte de dados obtenha acesso global à comunidade sem a necessidade de conhecer a estrutura dos padrões de metadados.

- **Linguagem de Representação**

Para resolver o problema da heterogeneidade é necessário definir uma forma de representação comum capaz de prover o diálogo entre o mediador e as fontes. Neste trabalho foi utilizada a linguagem de representação XML, devido, principalmente, à sua flexibilidade para representar dados estruturados e semi-estruturados e à facilidade de conversão de dados para este formato. Sendo assim, a tecnologia XML

está presente na descrição dos modelos de correspondência (mapeamentos) e nas sub-consultas enviadas pelo mediador às fontes de dados. A linguagem de consulta XML utilizada foi *XPath* [83]. O padrão *XPath*, proposto pelo W3C, é uma linguagem que permite referenciar partes de um documento XML.

- **Protocolo de comunicação**

O protocolo de comunicação utilizado para envio de consultas e recebimento de dados dos tradutores é o SOAP (*Simple Object Access Protocol*) [84] [85]. O SOAP é um protocolo que define um caminho de comunicação entre aplicações em diferentes domínios. Basicamente, encapsula uma chamada de função em XML e retorna o resultado da execução [86]. Os dados passados como parâmetro e os valores retornados também são estruturados em XML.

A figura 5.2 apresenta os principais componentes da camada de mediação, os quais são descritos a seguir:

O Mediador Central recebe a consulta global (segundo o esquema comum) submetida pelo usuário e a reescreve em  $N$  sub-consultas. A reescrita corresponde a uma filtragem na consulta global, obtendo apenas os predicados genéricos que possuem correspondência num padrão (ou meta-esquema) de metadados específico  $i$  (onde  $i = 1.. N$ ). Desta forma, a sub-consulta reescrita é enviada para um Mapeador específico do padrão  $i$ .

O Mapeador  $i$  funciona como uma espécie de tradutor do modelo global para o padrão  $i$  (por exemplo, ARIADNE), convertendo a consulta com base nas regras de mapeamento definidas através de ontologias, de modo que esta consulta possa ser resolvida pelas fontes que implementam este padrão.

O Mediador  $i$  recebe a sub-consulta mapeada para o padrão  $i$  e a distribui para as fontes que implementam este padrão. Baseado num conjunto de regras, o mediador pode restringir a distribuição da sub-consulta para algumas fontes, bem como gerar consultas mais apuradas. Esta seleção de fontes aplica-se, por exemplo, quando uma sub-consulta requer LOs de língua portuguesa e uma determinada fonte contém apenas LOs de língua inglesa e, portanto, a consulta não seria submetida para esta fonte.

A introdução de uma subcamada de mediação surgiu como forma de organização e distribuição da carga de processamento, utilizando uma política de agrupamento de fontes de dados. Além disso permite que os sítios participantes da comunidade de *e-learning* conheçam apenas o padrão de metadados adotado, não sendo necessário conhecer o modelo global, reduzindo assim a complexidade na implementação dos tradutores.

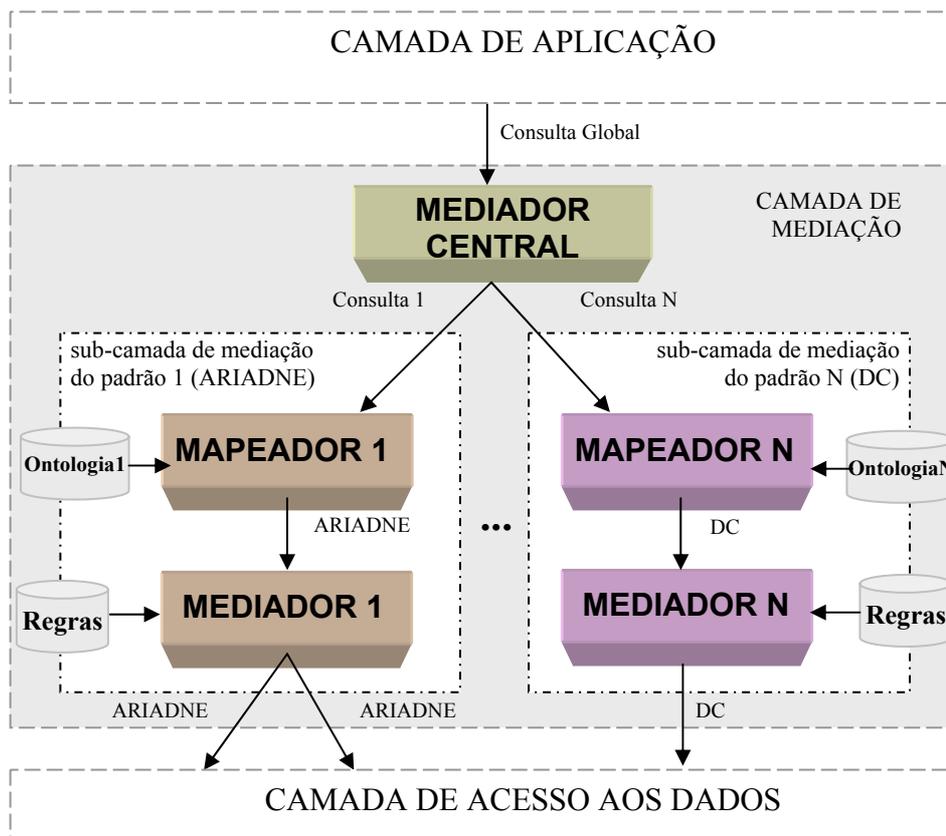


Figura 5.2: Camada de Mediação

Nas subseções subseqüentes são detalhados os serviços que provêm as funcionalidades dos componentes descritos na Figura 5.2.

### 5.2.2.1. Serviço de Registro das Fontes de Dados

As fontes de dados são registradas no LORIS através deste serviço, que gerencia informações de cada fonte de dados, tais como: dados de identificação, parâmetros de comunicação, etc.

Após a execução deste serviço é necessário cadastrar um conjunto de informações sobre a fonte através do serviço de gerência de informações, explicado na seção seguinte.

### 5.2.2.2. Serviço de Gerência de Informações

Este serviço é responsável por gerenciar informações relevantes a respeito das fontes, tais como: ontologias dos padrões e dos mapeamentos em OWL, descrições dos serviços dos tradutores (WSDL [87]), regras para refinamento de consultas, etc.

Essas informações devem manter-se atualizadas de forma a garantir o sucesso em todo o processo de mediação.

### 5.2.2.3. Serviço de Mediação

O serviço de mediação é responsável por fazer a interação entre os componentes envolvidos na camada mediação. Essa interação envolve três serviços, conforme apresentado na Figura 5.3:

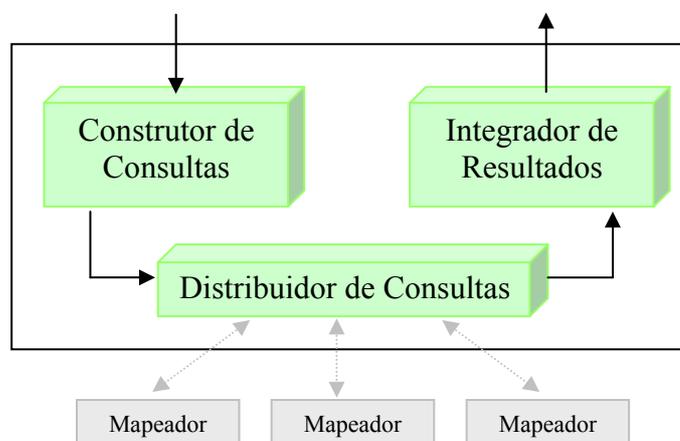


Figura 5.3: Serviços do Mediator Central

- Construtor de Consultas – recebe a consulta submetida pelo usuário e a decompõe em sub-consultas de acordo com a estrutura de cada padrão (ou meta-esquema) de metadados.
- Distribuidor de Consultas – envia as sub-consultas para seus respectivos Mapeadores, recebendo o resultado retornado por cada um e repassando-os para o Integrador de Resultados.
- Integrador de Resultados – responsável pela integração dos dados provenientes das fontes.

#### **5.2.2.4. Serviço Mapeador**

O serviço Mapeador é responsável por gerenciar o mapeamento entre diferentes esquemas de padrões de metadados. Deste modo, este serviço realiza o mapeamento de uma sub-consulta definida segundo o modelo global para uma sub-consulta definida segundo um determinado padrão (ou meta-esquema) de metadados utilizado nas fontes de dados. Todas as consultas requeridas pelo usuário são construídas no formato do modelo global e o serviço Mapeador é responsável em transformar esta consulta em sub-consultas baseadas em Ariadne, DC, LOM e qualquer outro padrão (ou meta-esquema) que venha ser incluído na arquitetura.

A definição do modelo de mapeamento é uma tarefa que exige um "entendimento semântico" entre os padrões, conforme descrito no capítulo 4.

#### **5.2.2.5. Serviço de Refinamento**

O serviço de Refinamento é um serviço do componente Mediador i. Este serviço é invocado pelo serviço Mapeador, com a finalidade de restringir a submissão da consulta a um subconjunto de fontes e/ou refinar as sub-consultas a serem enviadas às fontes de dados, de acordo com regras previamente definidas.

Uma regra aplica-se, por exemplo, quando uma sub-consulta requer LOs para uma faixa etária específica (adulto) e uma determinada fonte contém apenas LOs para outra faixa etária (criança). Neste caso, a consulta não seria submetida a

esta fonte. Para que isto aconteça, é necessário que a fonte em questão seja registrada previamente junto ao LORIS indicando a regra que especifica que a mesma possui apenas dados de uma determinada faixa etária (criança 0 a 12 anos). Uma regra é definida por uma fonte quando uma determinada condição é válida para todos os seus LOs.

No caso exemplificado, a fonte em questão sabe previamente que todos os seus LOs são voltados para a faixa etária de 0 a 12 anos, ou seja, sua base de LOs satisfaz a seguinte expressão (considerando uma fonte  $F_1$ ):

$$\{\forall x \mid x \text{ in } F_1, \mathbf{x.typicalAgeRange = "0-12"}\}$$

Considerando a regra definida pela expressão anterior, temos que a fonte  $F_1$  é descartada pela sub-consulta que busca LOs para a faixa etária adulta.

Além da restrição de fontes, o serviço também provê regras que possibilitam o refinamento das sub-consultas. Considere, por exemplo, a sub-consulta que deseja obter os LOs para a faixa etária adulta **ou** que possuam duração de dois meses. Aplicando esta consulta à fonte  $F_1$  (mesma fonte de exemplo anterior), o serviço de refinamento reescreve a consulta de modo que a mesma seja submetida à  $F_1$  contendo apenas o predicado "duração igual a dois meses", ou seja, eliminando o predicado da "faixa etária adulta", que conforme discutido anteriormente, não teria resultados nesta fonte.

### **5.2.3. Camada de Acesso aos Dados**

A camada de acesso aos dados é responsável pela interface entre os componentes de mediação e as bases de dados operacionais. Esta camada consiste de tradutores que são responsáveis por se comunicarem com diversos sistemas de gerência de banco de dados (SGBDs) assim como sistemas de arquivos.

#### **5.2.3.1. Serviço de Tradução**

O serviço de Tradução é responsável por fazer a interação entre os componentes envolvidos na camada de Acesso aos Dados. Essa interação é composta por dois serviços, conforme apresentado na Figura 5.4:

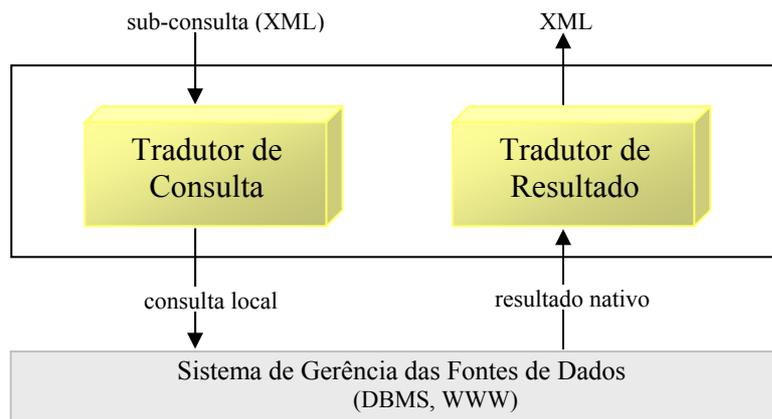


Figura 5.4: Serviços de Tradução

- Tradutor de Consulta – recebe a sub-consulta do mediador, a reescreve no modelo de dados utilizado pela fonte de dados e a submete para o sistema de gerência da fonte de dados local.
- Tradutor de Resultado – recebe, do sistema de gerência da fonte, o resultado em seu formato proprietário, traduzindo-o para o formato do modelo global.

### 5.3. Diagrama de Casos de Uso

Nesta seção é apresentado o diagrama de casos de uso referente aos serviços propostos pela arquitetura na seção anterior. O diagrama, além de oferecer uma visão geral das funcionalidades, configura-se como objeto de documentação do sistema, em conformidade com a linguagem UML (*Unified Modeling Language*) [88].

A figura 5.5 ilustra o diagrama de casos de uso do sistema LORIS, especificando a pesquisa por LOs, foco principal deste trabalho, assim como os casos de uso relativos à administração do ambiente.

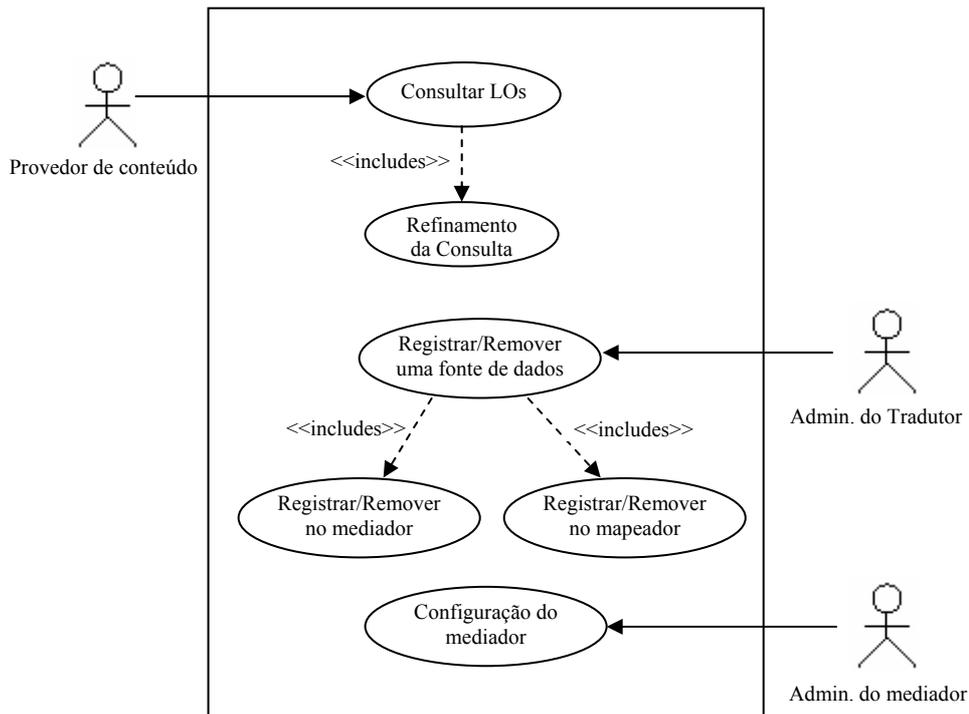


Figura 5.5: Diagrama de Casos de Uso

### **Consultar LOs**

O principal caso de uso do sistema LORIS consiste na pesquisa por recursos de *e-learning* (LOs). A consulta é submetida pelo usuário através de uma interface gráfica. O sistema opera sobre as fontes integradas, gerenciando a entrega das sub-consultas e a composição dos resultados obtidos.

### **Refinamento das Consultas**

Este caso de uso tem a finalidade de restringir a consulta a um subconjunto de fontes e/ou reescrever as sub-consultas a serem enviadas às fontes de dados, de acordo com regras previamente definidas.

### **Registrar/Remover uma Fonte de Dados**

A fonte de dados e seu respectivo tradutor são integrados ao LORIS através de um registro de informações relevantes para o funcionamento do sistema, descritos nas seções 5.2.2.1 e 5.2.2.2. Desta forma, a fonte de dados pode ser considerada no processamento da consulta. A fonte de dados também pode ser removida do sistema, o que implica na exclusão do registro.

### **Configuração do Mediador**

Consiste na configuração inicial do sistema, onde são definidos os padrões (ou meta-esquemas) de metadados adotados pelo LORIS, além da definição do modelo global construído a partir das ontologias dos padrões. Sendo assim, o administrador do mediador não só especifica os modelos como também as regras de correspondência entre eles.

Os casos de uso descritos anteriormente operam em conjunto, com o intuito de prover operações essenciais de busca, onde um interessado em LOs (uma pessoa ou uma aplicação de *software*) pode invocar um método de pesquisa; consultar uma rede de repositórios de LOs usando aquele método; obter e exibir um conjunto de LOs; visualizar os metadados correspondentes e selecionar um ou mais objetos (recurso digital) cujos endereços são apresentados como *hyperlinks*.

A seguir, descrevemos algumas possíveis consultas que podem ser resolvidas pelo sistema LORIS. Essas consultas exploram os metadados representados nos padrões abordados neste trabalho e auxiliam o usuário (professor, instrucionista, palestrante, alunos, etc.) a encontrar os LOs de seu interesse.

1. Encontre LOs que possam ser aplicados como questionário sobre o tema "Normalização".
2. Busque por tópicos que fazem parte do curso "Banco de Dados".
3. Encontre LOs que possam ser aplicados com tempo de aprendizagem inferior a 30 minutos.
4. Encontre figuras que representam o "Modelo Entidade-Relacionamento".
5. Busque por LOs direcionados a aprendizes cuja idade está entre "0-5" anos.
6. Pesquise LOs que possam ser exibidos no navegador "Netscape", versão "6.0".
7. Recupere os LOs que tenham como palavra-chave "SQL" e "álgebra relacional".

#### **5.4. Considerações Finais**

Neste capítulo foi apresentada uma arquitetura para integração de dados baseada em mediadores e tradutores como solução para o problema de integração de metadados de comunidades de *e-learning*. Estas comunidades apresentam ambientes heterogêneos e distribuídos em que informações encontram-se distribuídas em fontes de dados na *Web*. A arquitetura foi descrita em camadas, as quais foram subdivididas em componentes desenvolvidos como serviços *Web*. Também foram definidos os casos de uso do LORIS referente aos serviços propostos na arquitetura.