

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

ELISABETE TOMOMI KOWATA

**Metadados de Bancos de Dados
Relacionais: Extração e Exposição com
o Protocolo OAI-PMH**

Goiânia
2011

ELISABETE TOMOMI KOWATA

Metadados de Bancos de Dados Relacionais: Extração e Exposição com o Protocolo OAI-PMH

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Sistemas de Informação.

Orientador: Prof. Dr. João Carlos da Silva

Goiânia
2011

ELISABETE TOMOMI KOWATA

Metadados de Bancos de Dados Relacionais: Extração e Exposição com o Protocolo OAI-PMH

Dissertação defendida no Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Ciência da Computação, aprovada em 11 de novembro de 2011, pela Banca Examinadora constituída pelos professores:

Prof. Dr. João Carlos da Silva
Instituto de Informática – UFG
Presidente da Banca

Prof. Dr. Sean Wolfgang Matsui Siqueira
Departamento de Informática Aplicada – UNIRIO

Prof. Dr. Auri Marcelo Rizzo Vincenzi
Instituto de Informática – UFG

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Elisabete Tomomi Kowata

Graduou-se em Tecnologia em Processamento de Dados na Faculdade de Ciências Econômicas de Anápolis (FACEA) e em Licenciatura Específica em Matemática na Universidade Estadual Vale do Acaraú (UVA). Obteve os títulos de especialista em Redes de Computadores pela Universidade Estadual de Anápolis (UNIANA), Especialista em Agentes de Inovação e Difusão Tecnológica pelo Centro Federal de Educação Tecnológica de Goiás (CEFET-GO), Especialista em Gestão Financeira pelo Centro Universitário de Anápolis (UniEvangélica) e Especialista em Assessoria Linguística e Revisão Textual pela Universidade Estadual de Goiás (UEG). Durante o Mestrado, na Universidade Federal de Goiás (UFG), foi bolsista da Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG). Atua como Coordenadora do Núcleo de Inovação Tecnológica e *webdesigner* e também como professora dos Cursos de Graduação da Unidade Universitária de Ciências Socioeconômicas e Humanas de Anápolis (UnUCSEH/UEG).

À Deus pela vida e aos familiares e amigos pelo incentivo.

Agradecimentos

Agradeço a Deus pela vida e saúde.

Ao professor Dr. João Carlos da Silva pela orientação, dedicação, paciência, exigência e disponibilidade durante todo o período do mestrado e também por acreditar em meu potencial. Exemplo que seguirei por meu percurso profissional pelo profissionalismo e dedicação à profissão docente. Agradeço-lhe a confiança depositada em mim e as oportunidades de aprendizagem proporcionadas durante o curso.

Ao professor Dr. Cedric Luiz de Carvalho pelas sugestões que muito contribuíram para o desenvolvimento deste trabalho.

Ao professor Dr. Auri Marcelo Rizzo Vincenzi pelas correções e sugestões feitas nos artigos submetidos aos Congressos da área que muito contribuíram para a melhoria destes textos.

Aos professores Dra. Diane Castonguay, Dr. Hugo Alexandre Dantas do Nascimento, Dr. Humberto Longo, Dr. Claudio Nogueira de Meneses e Dr. Leandro Luís Galdino de Oliveira que me auxiliaram durante o curso.

A todos os funcionários, especialmente ao Sr. Edir que me acompanhou desde o início, aos amigos e demais professores do Instituto de Informática da Universidade Federal de Goiás.

Aos colegas do curso pela convivência ao longo destes meses e pelas novas amizades feitas. Um abraço especial aos colegas Valdemar, Luis Loja, André, Rogério, Adriana, Carine, Fabiana, Patrícia, Elisângela, Victor, Glauber, Renan, Raí, Jean, Halley, Cássio, Luciana, Leonel, Wandeir, Renata, Noeli, Marcelo Quinta, Sofia Larissa...

Ao Renan Rodrigues de Oliveira que me ajudou e me incentivou durante a pesquisa.

Ao professor Ms. Nelson de Abreu Júnior, ex-diretor da Unidade Universitária de Ciências Socioeconômicas e Humanas, pela amizade, compreensão e confiança, por ter sido o meu grande incentivador durante o curso.

Ao professor Ms. Marcelo Jose Moreira, diretor da Unidade Universitária de Ciências Socioeconômicas e Humanas, pela amizade, compreensão e confiança.

Aos colegas de trabalho, aos professores da Unidade Universitária de Ciências Socioeconômicas e Humanas de Anápolis e aos colegas da Administração Superior da

Universidade Estadual de Goiás que acompanharam a minha luta.

À professora Ms. Maria Elena Fregonesi Spínola e ao Dr. Francisco Spínola (Dr. Francis) pelos incentivos e por acreditarem em minha capacidade.

Ao grupo de pesquisa REDUCATIVA da Universidade Estadual de Goiás na pessoa da professora Dra. Mirza Seabra Toschi com quem aprendi muito no período em que participei do grupo de pesquisa *Leitura na Tela*.

Aos meus familiares pela formação e estabilidade que sempre me deram.

A FAPEG pelo apoio financeiro.

A todos que direta ou indiretamente me ajudaram a concretizar mais esta etapa da minha vida.

E principalmente a Deus pela perseverança que me deste e tem me dado em todos os dias de minha vida. Aprendi a perder para ganhar futuramente.

Culto é aquele que sabe onde encontrar aquilo que não sabe.

George Simmel,
Sociólogo alemão, 1858 - 1918.

Resumo

Kowata, Elisabete Tomomi. **Metadados de Bancos de Dados Relacionais: Extração e Exposição com o Protocolo OAI-PMH**. Goiânia, 2011. 125p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

Informações sobre um determinado assunto podem estar armazenadas em diferentes repositórios como banco de dados, bibliotecas digitais, planilhas eletrônicas, arquivos textos, páginas na *web* etc. Nesse contexto de fontes de dados heterogêneas, consultar, possivelmente em linguagem natural, integrar informações e promover interoperabilidade são tarefas que dependem, dentre outros fatores, do conhecimento prévio que um usuário tem sobre a localização, o proprietário, a descrição do conteúdo de cada fonte de informação. Mais especificamente, no caso de bancos de dados, essas informações não são, em geral, armazenadas no catálogo de um sistema gerenciador de bancos de dados; para obtê-las é necessário recorrer ao conhecimento do administrador desse banco. Outro fator que evidencia essa dependência é a ausência de mecanismos de busca a bancos de dados na *web* que acessam e tornam disponíveis as informações contidas nesses repositórios, devido ao fato desses dados estarem limitados às próprias organizações. Em um ambiente de compartilhamento de informações, é altamente relevante tornar possível o acesso aos metadados que descrevem uma fonte de dados, independentemente do meio e do formato em que esteja armazenada. Este trabalho tem como objetivo descrever um mecanismo para promover interoperabilidade de bancos de dados relacionais com outras fontes de informações, por meio da extração e exposição dos metadados usando o protocolo OAI-PMH.

Palavras-chave

Bancos de Dados Relacionais. Metadados. Interoperabilidade. Dublin Core. OAI-PMH.

Abstract

Kowata, Elisabete Tomomi. **Metadata of Relacional Database: Extraction and Exposition With OAI-PMH Protocol**. Goiânia, 2011. 125p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

Information about a particular subject can be stored in different repositories such as databases, digital libraries, spreadsheets, text files, web pages etc. In this context of heterogeneous data sources, to query, possibly in natural language, to integrate information and to promote interoperability are tasks that depend, among other factors, on the prior knowledge that an user has regarding location, owner, content description of each information source etc. More specifically, in the case of database, this information are not usually stored in a catalogue of the database management system and to obtain is necessary to resort to the administrator's knowledge database. Another factor is the absence of search engines to databases in the web that access and make available the information in those repositories, data are limited due to the organizations themselves. In a shared information environment, it is highly relevant to make possible access to metadata that describe a data source, regardlessly of the device and format in which is stored. This study aims to describe a mechanism to promote interoperability of relational databases with other sources of information through the extraction and exposing of metadata using OAI-PMH.

Keywords

Relational Database. Metadata. Interoperability. Dublin Core. OAI-PMH.

Sumário

Lista de Figuras	12
Lista de Tabelas	14
Lista de Códigos de Programas	15
Lista de Abreviaturas e Siglas	16
1 Introdução	17
1.1 Contexto e Motivação	17
1.2 Objetivos	19
1.3 Principais Problemas e Necessidades	20
1.4 Interoperabilidade	20
1.5 Metadados	21
1.6 Metadados no Contexto de Banco de Dados	24
1.7 Organização do Texto	26
2 Fundamentação Teórica	28
2.1 Trabalhos Relacionados	28
2.2 Catálogo de Sistemas de Bancos de Dados Relacionais	32
2.2.1 Catálogo no <i>DB2</i>	33
2.2.2 Catálogo no <i>MySQL</i>	33
2.2.3 Catálogo no <i>Oracle</i>	35
2.2.4 Catálogo no <i>PostgreSQL</i>	36
2.3 <i>Open Archives Initiative</i>	37
2.4 O Protocolo OAI-PMH	39
2.4.1 Objetivo do Protocolo OAI-PMH	41
2.4.2 Etapas para uma Coleta Eficiente de Metadados	42
2.5 Repositórios Digitais	46
2.6 <i>Dublin Core</i>	48
2.6.1 Objetivos do <i>Dublin Core</i>	49
2.6.2 Elementos do <i>Dublin Core</i>	50
3 Extração de Metadados de Bancos de Dados Relacionais	53
3.1 Definição dos Elementos do <i>Dublin Core</i>	54
3.1.1 Elementos <i>Dublin Core</i> em Bibliotecas Digitais	54
3.1.2 Significado dos Elementos do <i>Dublin Core</i> adotados pelas Bibliotecas Digitais	59
3.2 Utilizando Elementos <i>Dublin Core</i> para descrever Metadados de Bancos de Dados Relacionais	61

3.2.1	Exemplo de Uso da Estrutura Proposta	64
3.2.2	Um Mecanismo para Interoperabilidade de Banco de Dados com outras Fontes de Informação	65
4	Um Sistema de Extração e Exposição dos Metadados de Banco de Dados Relacionais	69
4.1	Objetivos	69
4.2	Principais Problemas, Necessidades e Escopo	70
4.3	Principais Atores	71
4.4	Requisitos de <i>Software</i>	72
4.5	Implementação do Protótipo	73
4.5.1	Tecnologias Empregadas	73
4.5.2	Estrutura da Tabela de Metadados para Exposição (TME)	74
4.5.3	Modelo de Solução Proposto	75
4.5.4	Visões dos Usuários do Sistema	77
4.5.5	Arquitetura do Sistema	78
4.5.6	Descrição do Protótipo	80
5	Testes/ Exemplos	93
5.1	Configuração do Ambiente de Teste	93
5.2	Exemplos de consultas	100
5.2.1	Coleta de Metadados	100
5.2.2	Sistema de Busca	107
6	Conclusão	109
6.1	Contribuições	110
6.2	Trabalhos Futuros	112
6.3	Produção Bibliográfica	113
	Referências Bibliográficas	114
A	Guia de Instalação	120
A.1	Tecnologias Empregadas	120
A.2	Configuração do Ambiente	120
A.3	Configuração do Banco de Dados	120
A.4	Instalação e Configuração do Aplicativo	121
A.5	Definição das Estruturas das Tabelas em SQL	122
B	Diagrama de Contexto e Diagrama de Entidade Relacionamento	124
B.1	Diagrama de Contexto do Sistema	124
B.2	Diagrama de Entidade Relacionamento	125

Lista de Figuras

1.1	Configuração de um sistema de banco de dados simplificado [20]	25
1.2	Metadados em data warehouse, segundo Inmon [31]	25
1.3	Gerenciamento de metadados em um sistema gerenciador de banco de dados relacional [37]	26
2.1	Dicionário de dados do banco de dados MySQL: INFORMATION_SCHEMA [55]	34
2.2	Modelo conceitual do INFORMATION_SCHEMA do banco de dados MySQL [47]	36
2.3	Metadata harvesting model [28]	40
2.4	Modelo de busca federada e de coleta de metadados/ <i>harvesting</i> (SULEMAN, 2002, p. 15) apud Ferreira [25]	42
2.5	Funcionamento básico do controle de fluxo do OAI-PMH [8]	44
2.6	Fluxo de execução de um repositório digital. Adaptado de [35]	47
3.1	Número de ocorrências dos elementos que aparecem na visualização simplificada	57
3.2	Número de ocorrências dos elementos que aparecem na visualização completa	58
3.3	Modelo do extrator de metadados proposto	66
3.4	Passos para a execução do mecanismo proposto	66
4.1	Classificação dos atores quanto ao uso da solução do sistema	72
4.2	Fluxo de extração e exposição de metadados de bancos de dados relacionais. Adaptado de [35]	76
4.3	Modelo de solução do sistema proposto	77
4.4	Arquitetura do sistema proposto de extração de metadados (METADATA-BASE EXTRACTOR)	79
4.5	Arquitetura de conteúdo adotada para o protótipo desenvolvido. Adaptado de [58]	80
4.6	Página inicial da aplicação	81
4.7	Interfaces de coleta de metadados	82
4.8	Link para consulta a conteúdos de banco de dados	84
4.9	Sistema de busca	84
4.10	Interface de consulta às informações de banco de dados	85
4.11	Página do administrador da aplicação	86
4.12	Página de inclusão de metadados	87
4.13	Interface do Repository Explorer [61]	90
4.14	Bancos de dados configurados no servidor local	91

4.15	Exemplo: relação de tabelas do banco de dados INFORMATION_SCHEMA	91
4.16	Exemplo: relação de campos da tabela TABLES	92
4.17	Exemplo: dados da tabela COLLATIONS	92
5.1	Interface para coleta de metadados: parâmetros de Identify	100
5.2	Resultado da coleta de metadados: Identify	101
5.3	Interface para coleta de metadados: parâmetros de ListMetadataFormats	101
5.4	Resultado da coleta de metadados: ListMetadataFormats	102
5.5	Interface para coleta de metadados: parâmetros de ListSets	103
5.6	Resultado da coleta de metadados: ListSets	103
5.7	Interface para coleta de metadados: parâmetros de ListIdentifier	104
5.8	Resultado da coleta de metadados: ListIdentifier	104
5.9	Interface para coleta de metadados: parâmetros de pesquisa por título ou palavras-chave ou assunto	105
5.10	Resultado da coleta de metadados: pesquisa por título ou palavras-chave ou assunto	105
5.11	Interface para coleta de metadados: parâmetros de ListRecords	106
5.12	Resultado da coleta de metadados: ListRecords	106
5.13	Interface para coleta de metadados: parâmetros de GetRecord	107
5.14	Resultado da coleta de metadados: GetRecord	107
5.15	Resultado de busca: listar repositórios	108

Lista de Tabelas

1.1	Exemplo de um registro simples de metadados	24
2.1	Comparativo das pesquisas mencionadas	31
2.2	Os verbos e seus argumentos	41
2.3	Elementos do Dublin Core [26]	51
3.1	Bibliotecas digitais pesquisadas	55
3.2	Exemplos de metadados de duas bibliotecas digitais pesquisadas	56
3.3	Atributos da tabela de metadados para exposição (TME)	61
3.4	Exemplo de uso da estrutura dos elementos do Dublin Core e elementos adicionais para descrição de metadados de um banco de dados fictício	64
4.1	Requisitos do software	72
4.2	Estrutura dos elementos do Dublin Core e elementos adicionais para descrição dos metadados de banco de dados	74
5.1	Descrição dos bancos de dados utilizados para teste	93
5.2	Bibliotecas digitais utilizadas para teste	97

Lista de Códigos de Programas

2.1	– Código HTML - Codificação do <i>Dublin Core</i> como parte de uma página <i>web</i> usando o tag <META>	52
4.1	– Arquivo de metadados gerado no formato XML	89
5.1	– Código SQL - Código SQL da tabela <code>metadbx_tb_oai_records</code>	98
5.2	– Código SQL - Código SQL da tabela <code>metadbx_cad_admdatabase</code>	99
5.3	– Código SQL - Código SQL da tabela <code>metadbx_cad_relatorio_bancodados</code>	99
A.1	– Código SQL - Código SQL da tabela <code>metadbx_tb_oai_records</code>	122
A.2	– Código SQL - Código SQL da tabela <code>metadbx_cad_admdatabase</code>	123
A.3	– Código SQL - Código SQL da tabela <code>metadbx_cad_relatorio_bancodados</code>	123

Lista de Abreviaturas e Siglas

SIGLA	Descrição
API	Application Programming Interface
CSS	Cascading Style Sheets
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ISO	International Organization for Standardization
NISO	National Information Standards Organization
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
PHP	Hypertext Preprocessor
RME	Repositório de Metadados para Exposição
SGBD	Sistema Gerenciador de Banco de Dados
SQL	Structured Query Language
TME	Tabela de Metadados para Exposição
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WebApp	Aplicações baseadas na Web
WWW	World Wide Web
XML	Extensible Markup Language

Introdução

Não tema dar um grande passo, se for o indicado. Não se pode saltar sobre um abismo com dois pulos pequenos. David Lloyd George, estadista inglês, 1863-1945.

Este Capítulo descreve o contexto, a motivação, os objetivos e os principais problemas e necessidades identificados na pesquisa. São apresentados conceitos de interoperabilidade e metadados, além de contextualizar os metadados em bancos de dados. A Seção 1.1 fornece o contexto e motivação da pesquisa, seguida dos objetivos do trabalho que são expostos na Seção 1.2 e dos principais problemas e necessidades identificados (Seção 1.3). As Seções 1.4 e 1.5 trazem as definições de interoperabilidade e metadados, respectivamente. A Seção 1.6 apresenta os metadados no contexto de banco de dados e por fim, a Seção 1.7 descreve a organização do trabalho.

1.1 Contexto e Motivação

A *World Wide Web* (WWW) tornou-se um meio em que os usuários podem encontrar todas as informações de que necessitam, desde uma simples receita a uma complexa descrição de como funciona uma espaçonave ou de como se desenvolve pesquisa com células tronco. Essas pesquisas realizadas por mecanismos de buscas gerais, que passam continuamente por aperfeiçoamento para a recuperação de informação, geram alta revocação (grande quantidade de informação recuperada), porém baixa precisão (pouca informação relevante) [17]. Porém, grande parte dessas informações não são encontradas na *web* pública ou visível que se conhece e sim em espaço conhecido como "*deep web*"¹, *web* invisível ou *web* oculto. As informações armazenadas nesses espaços ficam fora do alcance público, porque ou são privados ou é preciso pagar para se ter acesso a elas, razão esta porque ainda são poucos exploradas. Segundo Gill et al. [28],

¹*web* secreto

Uma pesquisa publicada em 2001 declarou que a *web* secreta era quinhentas vezes maior que a *web* visível/indexada, embora atualmente muito pouca informação possa ser extraída a partir da *web* secreta; em termos de evolução da *Internet*, cinco anos é o equivalente à era geológica.

Típicos representantes desse espaço são os bancos de dados em que as informações não podem ser pesquisadas por mecanismos de buscas gerais como Google e Yahoo, por exemplo, já que os conteúdos desses repositórios não são indexados.

Recuperar informações de bancos de dados relacionais é uma atividade que exige conhecimento das estruturas de armazenamento e da sintaxe da linguagem de acesso, o que é um obstáculo aos usuários. Além disso, informações como uma descrição do conteúdo do banco de dados, a localização, o proprietário etc. não são, em geral, armazenadas no catálogo de um sistema gerenciador de banco de dados. Para obter essas informações é necessário recorrer aos conhecimentos do administrador desse banco, o que torna inviável aos usuários da *web*. O acesso restrito e a ausência de mecanismos de busca a banco de dados na *web* são fatores que precisam ser resolvidos para promover a interoperabilidade. Em se tratando de bancos de dados, o cenário não contribui favoravelmente para a interoperabilidade desse tipo de repositório com outros repositórios disponíveis na *web*, devido aos dados estarem limitados às próprias organizações.

A principal questão que deve ser considerada é que em um ambiente de compartilhamento de informações é altamente relevante tornar possível o acesso aos metadados que descrevem uma fonte de dados. Mas como tornar possível a pesquisa em bancos de dados, uma vez que não estão ao alcance dos mecanismos de busca e ficam restritos às organizações?

A partir dessa questão, constata-se que é altamente relevante promover a interoperabilidade do banco de dados relacional com outras fontes de informações, o que pode ser obtido por meio da extração e exposição dos metadados. O protocolo OAI para coleta de Metadados (OAI-PMH - *Open Archives Initiative-Protocol for Metadata Harvesting*) tornou-se um mecanismo popular para publicação e compartilhamento de coleções digitais [52]; e a Iniciativa de Arquivos Abertos desenvolve e promove padrões de interoperabilidade que visam facilitar a disseminação eficiente de conteúdos e têm raízes no movimento de acesso aberto e repositórios institucionais [27].

A razão da escolha de banco de dados relacional é principalmente devido à grande parte dos dados existentes nas organizações ser armazenada nesse tipo de repositório, e seria altamente relevante incluir essas fontes de informação em uma consulta realizada no contexto da *web*.

Nesse contexto em que as informações estão armazenadas em fontes de dados heterogêneas, o questionamento que se faz é: Como acessar informações armazenadas

em diversas fontes, possivelmente heterogêneas, para que seja possível a realização de consultas por diferentes usuários?

A primeira motivação da pesquisa tem relação com a administração dos bancos de dados existentes em uma organização, que adota o padrão de qualidade em tecnologia da informação possui mecanismo próprio de administração e manutenção de banco de dados, porém outras que não possuem esse mecanismo de qualidade podem se perder com os inúmeros bancos criados e os recursos ficarem dispersos devido à falta de administração. Além da administração dos bancos de dados, outro fator importante é o compartilhamento das informações de bancos de dados dentro da organização. A proposta desse trabalho permitirá que os bancos de dados criados sejam catalogados e compartilhados entre os setores da organização.

A segunda motivação está na formação de uma base de dados geral de informações sobre bancos de dados. O proprietário do banco que desejar compartilhar com demais usuários da *web* o banco de dados, bastará disponibilizar uma descrição do banco e oferecer mecanismo de pesquisa a essas informações. Esse passo permitirá que o banco de dados possa ser indexado e os dados compartilhados com demais usuários da grande rede.

Uma terceira motivação da pesquisa está em permitir que usuários realizem consultas em ambientes heterogêneos, tornando possível a pesquisa independentemente da fonte em que os dados estejam armazenados. O grande desafio é fazer com que essas "ilhas de informações" sejam integradas apesar de armazenadas em fontes heterogêneas.

Outra motivação está na ampliação das fontes de pesquisa. As pesquisas em bibliotecas digitais abarcam fontes como artigos, periódicos, dissertações, imagens e músicas. Com este trabalho, as bibliotecas digitais podem estender a pesquisa também a bancos de dados.

Este trabalho poderá auxiliar pesquisas na área de interoperabilidade de bancos de dados com outros repositórios da *web*, facilitando a experimentação de propostas para recuperação de informações de banco de dados relacionais.

Portanto, a importância do trabalho está na administração e gerenciamento dos bancos de dados de uma organização, na formação de uma base de dados geral de informações de bancos de dados, na ampliação das fontes de pesquisas, possibilidade dos usuários realizarem consultas a esses dados por meio de mecanismos de buscas existentes, além de contribuir com a recuperação de informação que se encontram na *web* secreto.

1.2 Objetivos

Ao considerar o contexto e a motivação apresentados na seção anterior e o objetivo geral descrito anteriormente, os objetivos específicos deste trabalho são:

- auxiliar pesquisas na área de interoperabilidade de bancos de dados com outros repositórios da *web* facilitando a experimentação de propostas para recuperação de informações de bancos de dados relacionais;
- contemplar uma abordagem comumente utilizada em bibliotecas digitais, o que assegura uma maior aplicabilidade para bancos de dados relacionais;
- facilitar a disponibilização e a coleta de dados;
- favorecer a interoperabilidade, sem contudo exigir alterações nas fontes de dados, neste caso, dos bancos de dados relacionais. Assim, não é preciso ter acesso aos bancos de dados para extrair metadados, já que os metadados serão fornecidos pelos administradores de bancos de dados;
- propor um método para descrever e extrair metadados de bancos de dados relacionais;
- propor um mecanismo de exposição de metadados de bancos de dados relacionais;
- apresentar os resultados obtidos com o desenvolvimento de um protótipo para demonstrar a viabilidade da proposta apresentada.

1.3 Principais Problemas e Necessidades

Os principais problemas levantados durante o desenvolvimento da pesquisa foram:

- Informações sobre um determinado assunto podem estar armazenadas em diferentes repositórios: banco de dados, bibliotecas digitais, páginas *web* etc.;
- Uma única consulta de usuário, possivelmente em linguagem natural, pode acarretar uma pesquisa em diferentes fontes de dados;
- Como favorecer a interoperabilidade de dados armazenados em bancos de dados com dados provenientes de outras fontes?

1.4 Interoperabilidade

Uma vez que o objetivo deste trabalho está em promover a interoperabilidade do banco de dados relacional com outras fontes de informações por meio da extração e exposição dos metadados usando o protocolo OAI-PMH, esta seção apresenta algumas definições sobre interoperabilidade e o contexto que este termo será empregado no decorrer do texto.

Segundo Miller [46] *apud* [17], o termo "interoperável" é a capacidade de operar em conjunto e o conceito é cada vez mais utilizado na gestão da informação.

Já para *International Organization for Standardization (ISO) apud* [17], Interoperabilidade é a "habilidade de dois ou mais sistemas (de computadores, meios de comunicação, redes, *software* e outros componentes de tecnologia da informação) de interagir e de intercambiar dados de acordo com um método definido, de forma a obter os resultados esperados".

A página da Wikipédia.org [59] conceitua interoperabilidade como a capacidade de um sistema (informatizado ou não) de se comunicar de forma transparente (ou o mais próximo disso) com outro sistema (semelhante ou não). Para um sistema ser considerado interoperável é muito importante que ele trabalhe com padrões abertos.

Em OAforum [51], interoperabilidade é a capacidade de sistemas, serviços e organizações trabalharem em conjunto para alcançarem objetivos comuns ou diversos. Na área técnica é apoiada por padrões abertos para comunicações entre sistemas e para descrições de fontes e coleções, entre outros. No OAI-PMH a interoperabilidade é considerada principalmente no contexto de descoberta e acesso a fontes. Este trabalho utilizará esta definição no decorrer do texto.

A interoperabilidade entre os repositórios digitais tem o objetivo de promover o acesso simultâneo aos dados contidos em diversos repositórios digitais, de forma a maximizar as buscas e reduzir o tempo de resposta [16].

Para [2] *apud* [17], o objetivo da interoperabilidade é desenvolver serviços coerentes para os usuários, a partir de recursos informacionais que são tecnicamente diferentes e gerenciados por diferentes organizações. Isto requer acordos de cooperação em três níveis: técnico, de conteúdo e organizacional. O nível técnico proporciona a interoperabilidade tecnológica, representada nas definições citadas. O nível de conteúdo remete a interoperabilidade semântica, em que a representação e organização do conhecimento são áreas chaves a serem estudados. O nível organizacional se refere a interoperabilidade política, quando organizações se reúnem com o intuito de alcançar a interoperabilidade, implementando padrões e tecnologias que cooperem com este objetivo. A interoperabilidade política viabiliza a gestão articulada entre as organizações e os sistemas de informação, ela "depende fundamentalmente" da criação de organizações detentoras de canais e fóruns adequados, nos quais a discussão e o consenso possam se estabelecer e as decisões possam ser tomadas endossadas pelo grau de representatividade dessas organizações [43].

Na próxima Seção são apresentados conceitos de metadados, que também é importante para o entendimento do trabalho.

1.5 Metadados

Na literatura sobre metadados, a definição mais utilizada é a de que metadados representam "dados sobre dados" (Baeza-Yates e Ribeiro-Neto [3], Gill et al. [28], Niso

Press [50], *University of Queensland Library* [65] dentre outros).

A definição que se encontra na página mantida pela *University of Queensland Library* [65] diz que

Metadados são dados estruturados que descrevem as características de uma fonte. Eles compartilham características muito similares para catalogar dados que são obtidos em bibliotecas, museus e arquivos. O termo "meta" deriva da palavra grega que significa "uma natureza de alta ordem ou tipo mais fundamental". O registro de metadados consiste de um número de elementos pré-definidos representando atributos específicos de uma fonte e cada elemento pode ter um ou mais valores.

Em Niso² Press [50], metadados são informações estruturadas que descrevem, explicam, localizam, ou seja, tornam fácil recuperar, usar ou gerenciar uma fonte de informação. Em Baeza-Yates e Ribeiro-Neto [3], metadados são informações na organização dos dados, os vários domínios de dados e a relação entre eles. Por exemplo, em um sistema gerenciador de banco de dados, o esquema especifica alguns dos metadados, isto é, o nome das relações, os campos ou atributos de cada relação, o domínio de cada atributo etc.

Formas comuns de metadados associados com texto incluem o autor, a data de publicação, a fonte da publicação, o tamanho do documento (em páginas, palavras, *bytes* etc.) e o gênero do documento (livro, artigo, memorando e outros). Por exemplo, em *the Dublin Core Metadata Element Set* [19] é proposto um conjunto de quinze campos para descrever um documento.

Outro tipo de metadado é caracterizado pelo problema do assunto que pode ser encontrado dentro do conteúdo do documento, referenciado como "Metadado Semântico". Segundo Madnick (apud Bonifacio, 2010) [4] "os metadados semânticos fornecem informações sobre o significado dos dados disponíveis e seus relacionamentos semânticos". Exemplo disso são dados que descrevem o conteúdo semântico de um valor de dado (como unidades de medida e escala), ou dados que fornecem informações adicionais sobre sua criação (algoritmo de cálculo ou derivação da fórmula usada), linhagem dos dados (fontes) e qualidade (atualidade e precisão). Para padronizar termos semânticos, muitas áreas usam ontologias específicas, que são taxonomias hierárquicas de termos que descrevem certos tópicos de conhecimento.

Os metadados podem ser classificados em três tipos [18]: (1) metadados descritivos descrevem uma fonte de informação para fins de identificação e recuperação utilizando elementos como título, autor, resumo e palavras-chave, (2) metadados estruturados descrevem a organização interna dos objetos e das relações entre eles; o exemplo mais comum é o esquema do banco de dados, e (3) metadados administrativos apoiam as ativida-

²National Information Standards Organization (NISO)

des de gerenciamento do acervo de recursos de informação como controle de permissões de acesso, localização de arquivos e critérios de avaliação da qualidade.

Metadados descritivos revelam parte da semântica das fontes de informação, como a intenção de uso, além da descrição ou resumo do seu conteúdo (SHETH 1998 *apud* SANTOS [18]), fornecendo o contexto para entendimento dos dados através do tempo. Contexto é um aspecto importante na semântica dos dados e ajuda os analistas a interpretarem o significado atribuído a estes dados (AGRAWAL et al. 2009 *apud* SANTOS [18]).

As informações sobre qualidade de dados associadas às fontes de informação são uma parte do grupo de elementos de metadados administrativos [18]. Qualidade de dados é um conceito multidimensional que representa a visão, os critérios e as métricas para avaliar, interpretar e melhorar as fontes de informação além de permitir a seleção das dimensões relevantes dentro de um contexto de aplicação (SATTLER 2008 *apud* SANTOS [18]). As medidas são qualitativas, baseadas no julgamento de especialistas e usuários, e quantitativas, considerando as dimensões de avaliação (BATINI e SCANNAPIECO 2006 *apud* SANTOS [18]).

Para aumentar a confiabilidade dos usuários em relação às fontes de informações consideradas relevantes, além do conhecimento quanto à qualidade das mesmas, é necessário ter ciência de que estas são governadas por processos adequados, são gerenciadas por seus tutores (*Information Steward*) e possuem controle de acesso adequado, e para isto estas informações também devem ser registradas como metadados (GODINEZ et al. 2010 *apud* SANTOS [18]).

De acordo com Godinez et al. (2010 *apud* SANTOS [18]) metadados também podem ser classificados em técnicos e de negócio. Metadados técnicos contém informações referente a localização, protocolos de acesso e esquema físico e lógico das fontes de dados. Enquanto que, a maioria das informações textuais sobre os dados que usuários de negócio acrescentam nos modelos de dados de alto nível (HOBBERMAN et al. 2009 *apud* SANTOS [18]) e fornecem esclarecimento contextual são consideradas metadados de negócio como, por exemplo, as definições dos conceitos no negócio, o responsável pela definição, indicadores de qualidade das informações, critérios de segurança, indicação dos produtores e dos consumidores da informação além de vocabulários e taxonomias usadas para definir a terminologia e hierarquia dos conceitos do negócio.

Metadados podem ser extraídos de diversas fontes como políticas e procedimentos organizacionais, modelos de processos de negócio, mas principalmente de pessoas, no papel de responsáveis pelas aplicações e processos que geram e manipulam as informações (GODINEZ et al. 2010 *apud* SANTOS [18]).

A Tabela 1.1 apresenta um exemplo de um registro simples de metadados.

Tabela 1.1: *Exemplo de um registro simples de metadados*

Nome do elemento	Valor
Título	Banco de Dados <i>PostgreSQL</i>
Criação	Elisabete Tomomi Kowata
Publicação	Universidade Federal de Goiás
Tipo	Text
Data	09/06/2010
Formato	Text/html
Identificador	http://www.ufg.br/sgbd/meta/mainmenu.html
Relação	Gerenciador de conteúdo Joomla - versão 1.6

1.6 Metadados no Contexto de Banco de Dados

Uma característica fundamental da abordagem de um banco de dados é que o sistema de banco de dados possui não apenas o banco de dados, mas também uma completa definição ou descrição da estrutura desse banco e de suas restrições. Essa definição está armazenada no catálogo do Sistema Gerenciador de Banco de Dados (SGBD), o qual contém informações como a estrutura de cada arquivo, o tipo e o formato de armazenamento de cada item de dado e várias restrições sobre eles. A informação armazenada no catálogo é chamada metadados e descreve a estrutura do banco de dados primário [20] representada na Figura 1.1.

Ainda no contexto de banco de dados, pode-se exemplificar o *Data Warehouse* (DW). Segundo Inmon [31], um *Data Warehouse* é um conjunto de dados orientado por assunto, integrado, variável com o tempo e não-volátil, que fornece suporte ao processo de tomada de decisão do negócio.

Um componente essencial de uma arquitetura de *Data Warehouse* é o metadado e as ferramentas para o seu gerenciamento e recuperação. Ralph Kimball [21] descreve metadados como o DNA do *Data Warehouse*, os quais definem os elementos e a forma com que eles trabalham juntos.

Um esquema básico de criação e atualização de um *Data Warehouse* aparece na Figura 1.2.

No esquema apresentado, um *Data Warehouse* (DW) é atualizado por meio de fontes externas e internas por extração, filtragem, transformação e carga. Informações sobre essas fontes são então armazenadas no repositório de metadados que inclui informações também do próprio DW e de *Data Marts* e aplicações que os acessam. A parte final fica para um conjunto de programas que fazem análises dos dados contidos no *Data Warehouse*. Os *Data Mart* são subconjuntos de dados de um *Data Warehouse* (armazém

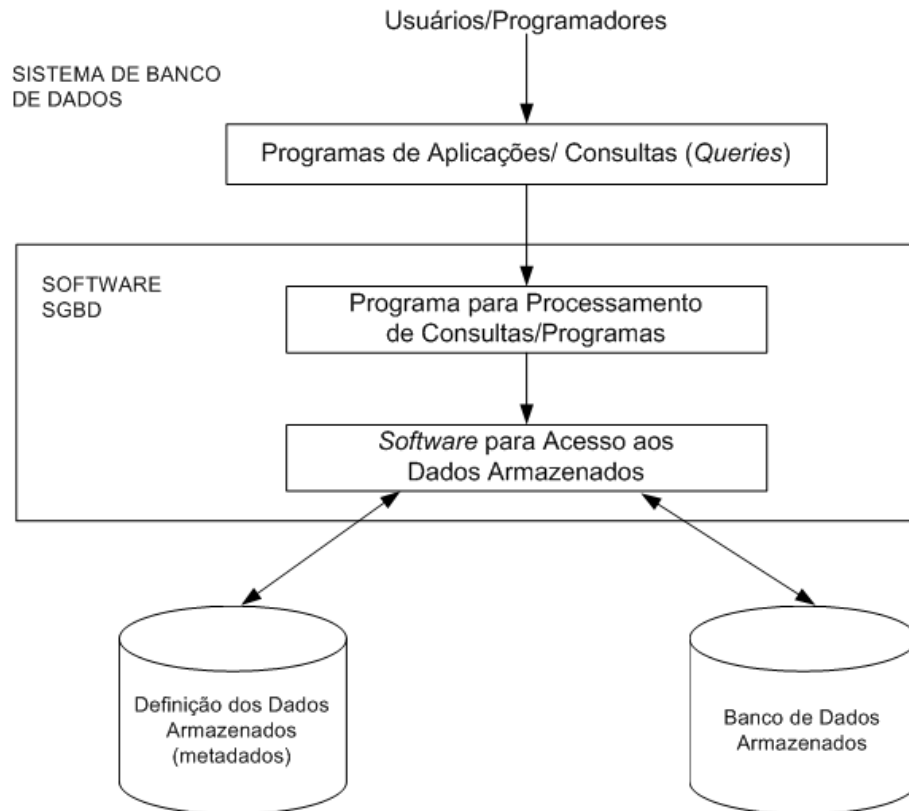


Figura 1.1: Configuração de um sistema de banco de dados simplificado [20]

de dados) e geralmente são dados referentes a um assunto em especial, como por exemplo estoque, compras ou recursos humanos, em diferentes níveis de sumarização como vendas e despesas mensais, compras quinzenais, que focalizam uma ou mais áreas específicas. Seus dados são obtidos do *Data Warehouse*, desnormalizados e indexados para apoiar intensa pesquisa. *Data Marts* extraem e ajustam porções de *Data Warehouses* aos requisitos específicos de grupos ou departamentos.



Figura 1.2: Metadados em data warehouse, segundo Inmon [31]

A Figura 1.3 mostra a diferenciação entre banco de dados, meta-banco de dados e as diferentes partes do meta-banco de dados; ilustra ainda como este é usado para processar as consultas dos usuários. O gerenciador de metadados é o núcleo do sistema de banco de dados e pode ser revisitado com a perspectiva de gerenciador de metadados, mais especificamente para aumentar a funcionalidade do banco de dados relacional.

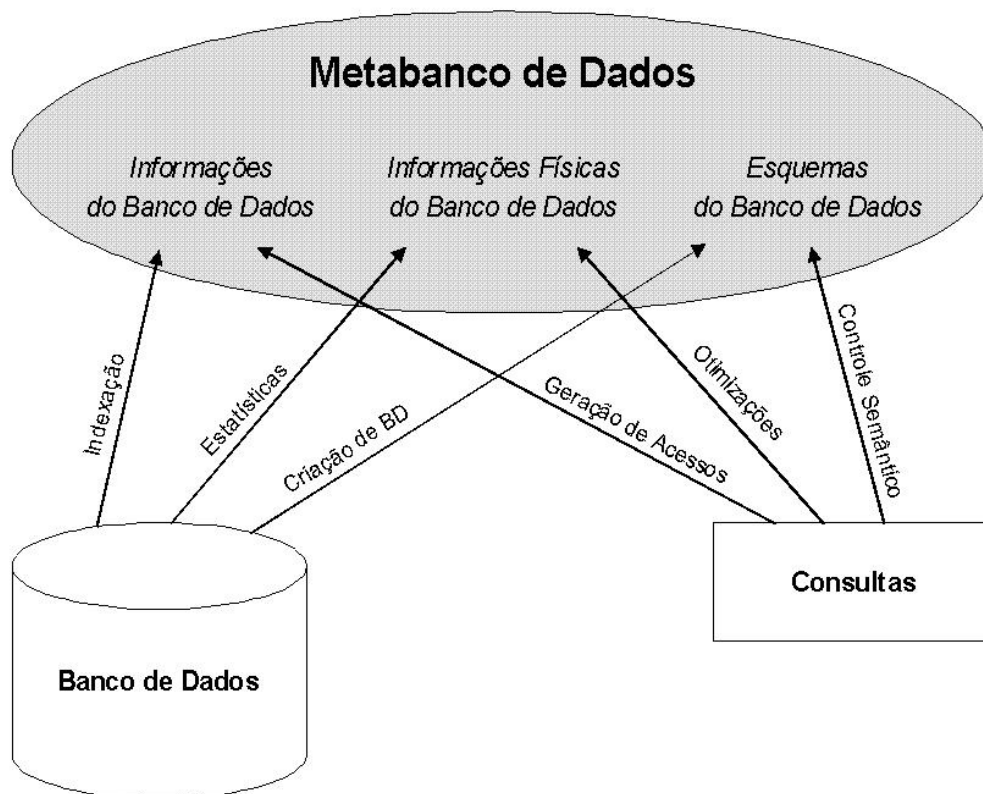


Figura 1.3: Gerenciamento de metadados em um sistema gerenciador de banco de dados relacional [37]

1.7 Organização do Texto

Além deste capítulo inicial que apresentou a introdução, o contexto em que o trabalho está inserido, a motivação e os objetivos deste estudo, o restante do texto está organizado conforme descrito nos próximos parágrafos.

No Capítulo 2 são apresentadas as fundamentações teóricas utilizadas para o desenvolvimento deste trabalho. Os principais assuntos apresentados são: catálogos de sistemas gerenciadores de bancos de dados, *Open Archives Initiative* e suas características, Protocolo OAI-PMH, repositórios digitais, objetivos e elementos do *Dublin Core* e conhecimento de algumas ferramentas que automatizam o processo de extração de metadados de documentos.

O Capítulo 3 apresenta uma proposta de um método de extração de metadados de bancos de dados relacionais. Nesta etapa foram levantadas as bibliotecas digitais e os elementos utilizados para descrever os conteúdos. A partir do levantamento foram definidos os elementos do *Dublin Core* (DC) e criados outros elementos adicionais para descrever os metadados de bancos de dados.

O Capítulo 4 apresenta uma solução do sistema de extração e exposição dos metadados de bancos de dados relacionais, bem como uma proposta de arquitetura do sistema.

O Capítulo 5 apresenta os resultados obtidos por meio da exposição de exemplos e recuperação dos metadados expostos.

O Capítulo 6 apresenta as conclusões finais, contribuições, trabalhos futuros e produção bibliográfica.

Com o objetivo de orientar a instalação do *software*, o apêndice A apresenta um "guia de instalação do aplicativo", que traz a relação de tecnologias utilizadas, bem como a orientação para configurar o ambiente, o banco de dados e o servidor *web* para a execução do aplicativo. O Apêndice B ilustra os diagramas de contexto e de entidade relacionamento do sistema.

Fundamentação Teórica

Se você construiu castelos no ar, não terá desperdiçado seu trabalho, pois no alto é onde devem estar. Agora coloque fundações embaixo deles. Henry David Thoreau, naturalista e escritor norte-americano, 1817-1862.

Este capítulo apresenta os trabalhos relacionados que tratam de extração e exposição de metadados com o protocolo OAI-PMH e também catálogos de alguns bancos de dados relacionais como o *MySQL*, *DB2*, *PostgreSQL* e o *Oracle*. O capítulo segue com a apresentação do Protocolo de Iniciativa de Arquivos Abertos (*Open Archives Initiative-OAI*), que possibilita a coleta de metadados expostos e descritos por meio do padrão *Dublin Core* e publicada na *web*. Além dos conceitos apresentados, conhecer as ferramentas que automatizam o processo de extração de metadados é importante para destacar as características que se deseja imprimir no protótipo a ser desenvolvido. Nesse sentido, algumas ferramentas que apoiam a extração de metadados também foram investigadas e descritas.

2.1 Trabalhos Relacionados

Nesta seção são apresentados alguns trabalhos relacionados que tratam da extração de metadados e sua disponibilização por meio do protocolo OAI-PMH.

Extração e exposição de metadados utilizando padrões como o *Dublin Core* e o protocolo OAI-PMH são utilizados no contexto das bibliotecas digitais e das fontes multimídias. Alguns trabalhos que tratam dessa temática são [6], [13], [9] e [15]. A criação e a adoção da OAI foram motivadas pelo fato de essa iniciativa facilitar a disseminação dos artigos científicos, ocasionando um aumento da oferta e uma rápida divulgação dos trabalhos desenvolvidos pela comunidade científica.

O trabalho de Cardoso [13] discute o crescimento do interesse na criação de bibliotecas digitais a partir do surgimento e da disseminação da *web*, que trouxe consigo a necessidade de ferramentas que facilitassem a publicação, o gerenciamento e a recuperação da informação digital. O autor levanta uma observação de uma ampla gama de

bibliotecas digitais, que se caracterizam como serviços que gerenciam e disponibilizam documentos digitais de forma mais estruturada do que convencionalmente se observa na *web*. Nesses serviços, em geral, os documentos são descritos com metadados, recuperados por meio de ferramentas de busca estruturada e visualizados em interfaces apropriadas. Cardoso aponta as limitações de grande parte das bibliotecas digitais existentes, dentre elas a ausência de mecanismos de integração de dados de maneira a fornecer ao usuário acesso unificado e transparente aos repositórios gerenciados por diferentes serviços. Esse problema é conhecido na literatura como o problema da interoperabilidade de bibliotecas digitais. Dentro desse contexto, o autor desenvolveu o Clio-i, uma arquitetura de sistema para gerenciamento de bibliotecas digitais. Essa arquitetura apresenta características desejáveis como um módulo de recuperação de documentos, um visualizador de documentos e um módulo de interoperabilidade entre repositórios digitais. Para prover o mecanismo de integração foi implementada no Clio-i uma extensão do protocolo OAI-PMH. O módulo integrador do Clio-i é composto de duas partes: (1) o Clio-i *Data Provider*, responsável por exportar os metadados dos documentos gerenciados localmente, de acordo com os padrões estabelecidos no OAI-PMH; e (2) o Clio-i *Service Provider*, que realiza a coleta de informações a partir de qualquer provedor de dados remoto baseado em OAI-PMH. O protótipo de sistema implementado foi validado em dois estudos de caso, operacionalizando centenas de milhares de registros e efetivando a extensão do protocolo OAI-PMH adotada.

Já em Contessa [9] é apresentado um serviço de geração de metadados compatível com o padrão OAI, desenvolvido para disponibilizar os metadados sobre os artigos publicados nas conferências ou periódicos que utilizem o *Journal and Event Management System* (JEMS). Os metadados gerados são disponibilizados para bibliotecas digitais como a Biblioteca Digital Brasileira de Computação (BDBComp). O serviço é um provedor de dados OAI, que utiliza dados provenientes do sistema JEMS e os disponibiliza no formato *Dublin Core* a bibliotecas digitais por meio do protocolo OAI-PMH, o qual é baseado em *Extensible Markup Language* (XML). O provedor de dados proposto pelo autor foi desenvolvido para uso com a biblioteca digital da Sociedade Brasileira de Computação (SBC) e oferece um aumento da quantidade e da qualidade dos metadados disponibilizados sobre os artigos publicados nos eventos periódicos gerenciados pelo JEMS.

O trabalho de De Oliveira et al. [16] aborda a recuperação contextualizada de documentos em um conjunto de bibliotecas digitais integrado por meio do protocolo OAI-PMH. Neste caso, apenas os documentos que fazem parte de um domínio especificado devem ser recuperados. No processo de contextualização das consultas são utilizadas ontologias e é feita a análise do conteúdo dos artigos da *Wikipédia* (por meio da manipulação de sua API (*Application Programming Interface*)), a qual é utilizada como um repositório

de conhecimento auxiliar. Essa é uma vertente no campo de recuperação de informações que tem se destacado principalmente com a explosão da *World Wide Web*.

Não foi encontrado um número significativo de referências com estudos específicos sobre a extração de metadados de banco de dados relacionais, porém alguns trabalhos isolados abordam a extração de metadados de banco de dados CDS/ISIS (*Computerized Documentation Systems/ Integrated Set of Information Systems*) em Kaloyanova et al. [36] e Jayakanth et al. [34].

O trabalho de Kaloyanova et al. [36] apresenta o CDS/SIS, um projeto realizado pela Organização da Agricultura e Alimentação das Nações Unidas (FAO) em colaboração com a Associação pela Documentação da Biblioteca e do Arquivo (DBA) da Itália. O CDS/ISIS é um sistema de recuperação de informação e armazenamento integrado da *United Nations Educational, Scientific and Cultural Organization* (Unesco), o qual é utilizado amplamente, especialmente na América Latina e na África. Existem centenas de aplicações baseadas no sistema CDS/ISIS as quais gerenciam referências bibliográficas, assegurando alta qualidade de conteúdo por meio do uso construído em arquivos de autoridades, orientações de entrada de dados e validações. Ainda permite exportar metadados em diferentes formatos. A metodologia de trabalho adotada inclui estudo, análise e avaliação de três soluções existentes para expor metadados a partir do repositório do banco de dados CDS/ISIS para o *framework* OAI.

Jayakanth et al. [34] apresentam uma abordagem dinâmica para fazer com que os bancos de dados legados, como CDI/ISIS, sejam interoperáveis com OAI e compatíveis com as bibliotecas digitais. Segundo Jayakanth et al. [34], existem duas possíveis abordagens para a interoperabilidade: estática e dinâmica. O trabalho discute a abordagem dinâmica para a execução do objetivo proposto por ser um meio simples para tornar interoperável os bancos de dados legados com outros padrões abertos.

O artigo de Haslhofer e Klas [30] aborda o problema de interoperabilidade de metadados quando se trata de acesso uniforme a objetos multimídia nos repositórios heterogêneos de mídia. Existem muitas técnicas de interoperabilidade que variam bastante o potencial para resolver as heterogeneidades estruturais e semânticas que podem existir entre os metadados armazenados em repositórios distintos. Os autores, além de proporcionarem uma visão geral do campo de interoperabilidade de metadados, que fornece uma categorização das técnicas de interoperabilidade existentes, descrevem suas características e comparam sua qualidade por meio da análise de seu potencial para resolver diversos tipos de heterogeneidades. Baseado nesse trabalho, profissionais da área e técnicos podem obter uma visão geral sobre a categorização de técnicas de interoperabilidade de metadados existentes e selecionar o método adequado para cenários específicos de integração de metadados. A análise demonstra que o mapeamento de metadados é a técnica apropriada em cenários de integração, em que a concordância sobre certo padrão de metadados não

é possível.

Librelotto et al. [40] apresentam o *Metamorphosis* - um ambiente orientado a *Topic Maps*, composto por módulos capazes de extrair dados de recursos heterogêneos de informação e disponibilizar navegação semântica sobre a informação extraída, usando, para isso, o paradigma de ontologias a fim de gerar uma visão homogênea desses recursos.

Sánchez et al. [63] descrevem *Voai* e *Xoai*, dois ambientes de *software* que facilitam a construção automática de servidores OAI para diversas coleções digitais. Os autores utilizam ferramentas disponíveis gratuitamente e não impõem exigência de programação por parte dos usuários.

A Tabela 2.1 apresenta uma síntese dos estudos descritos nesta seção. Cada trabalho foi avaliado em relação a cinco características: (1) exposição/extração de metadados; (2) uso do padrão *Dublin Core* para descrição de metadados; (3) uso do protocolo OAI-PMH; (4) interoperabilidade e (5) quanto ao objeto de estudo. A delimitação dos itens 2 e 3 para *Dublin Core* e para o protocolo OAI-PMH foi exclusivamente por este trabalho estar voltado para esse padrão e para esse protocolo, o que não impediria a ampliação das fontes de pesquisadas para outros padrões e protocolos.

A última linha da Tabela 2.1 refere-se à presente proposta de um novo artefato que contempla todas as características em que o objeto de estudo são os bancos de dados relacionais.

Tabela 2.1: Comparativo das pesquisas mencionadas

Autor(es)	Exposição e extração de metadados	Uso do padrão <i>Dublin Core</i> para descrição de metadados	Uso do protocolo OAI-PMH	Interoperabilidade	Objeto de estudo
Cardoso [13]	X	X	X	X	Arquitetura para gerenciamento de bibliotecas digitais
Contessa [9]	X	X	X	-	Disponibilizar os metadados sobre os artigos publicados nas conferências ou periódicos que utilizam o JEMS
De Oliveira et al. [16]	-	X	X	-	Recuperação contextualizada de documentos em um conjunto de bibliotecas digitais
Kaloyanova et al. [36]	X	-	X	-	Exposição de metadados a partir do repositório do banco de dados CDS/ISIS para o <i>framework</i> OAI
Jayakanth et al.					

[34]	-	-	-	X	Interoperabilidade dos bancos de dados legados como CDS/ISIS com OAI e compatível com as bibliotecas digitais
Haslhofer e Klas [30]	-	-	-	X	Interoperabilidade de metadados de objetos multimídia nos repositórios heterogêneos de mídia
Librelotto et al. [40]	-	-	-	X	Extração de dados de recursos heterogêneos
Sánchez et al. [63]	-	X	X	-	Descrição do <i>Voai</i> e <i>Xoai</i> , dois ambientes de <i>software</i> que facilitam a construção automática de servidores OAI para diversas coleções digitais
Kowata e Silva	X	X	X	X	Bancos de dados relacionais

Como pode ser observado na Tabela 2.1, a pesquisa que mais se aproxima desta proposta é a de Cardoso [13], que propõe uma arquitetura para gerenciamento de bibliotecas digitais. Da mesma forma que o trabalho de Cardoso [13] aponta a disseminação da *web*, evidencia a necessidade de ferramentas que facilitem a publicação, o gerenciamento e a recuperação da informação digital. Essa funcionalidade deveria ser estendida aos bancos de dados relacionais que se encontram dispersos pela *web*. Outra questão apontada pelo autor é a ausência de mecanismos de integração de dados, de maneira a fornecer ao usuário acesso unificado e transparente aos repositórios gerenciados por diferentes serviços. A partir do contexto apresentado pelo autor, desenha-se uma arquitetura que apresenta as características desejáveis em um sistema para gerenciamento de bibliotecas digitais: módulo de recuperação de documentos, visualizador de documentos e módulo de interoperabilidade entre repositórios digitais. Esta pesquisa busca, portanto, atender todas as características dos estudos anteriores direcionados aos bancos de dados relacionais.

Na Seção 2.2 são apresentados exemplos de catálogos de sistemas de bancos de dados relacionais como *DB2*, *MySQL*, *Oracle* e *PostgreSQL*.

2.2 Catálogo de Sistemas de Bancos de Dados Relacionais

Os catálogos de um sistema são os locais onde os sistemas gerenciadores de banco de dados relacionais armazenam os metadados do esquema, tais como informações sobre tabelas e colunas e informações de controle internos. Eles são, em sua maioria,

copiados do modelo do banco de dados durante a criação desse banco e, portanto, esses catálogos lhe são específicos. Uma característica fundamental da abordagem

O catálogo é usado tanto pelo *software* SGBD como pelos usuários do banco de dados que precisam de informações sobre a estrutura desse banco. Um pacote de *software* SGBD de propósito geral não está determinado para uma aplicação específica, portanto, para sua utilização será necessário acessar o catálogo para conhecer a estrutura dos arquivos no banco de dados, como o tipo e o formato dos dados que o programa acessará [20].

Algumas referências tratam o termo catálogo de banco de dados como metadados de banco de dados. Neste trabalho o termo será uniformizado para catálogo.

Nas subseções seguintes são apresentadas algumas características de catálogos de bancos de dados: *DB2*, *MySQL*, *Oracle* e *PostgreSQL*.

2.2.1 Catálogo no *DB2*

O catálogo no *DB2* [11] é um banco de dados de sistema que contém informações (descritores) relativas a diversos objetos que são de interesse do próprio sistema. Tabelas-base, visões, índices, banco de dados, planos de aplicação, direitos de acesso entre outros são exemplos desses objetos.

Uma vantagem significativa de um sistema relacional como o *DB2* é que o catálogo num sistema propriamente dito consiste de relações (ou tabelas de sistema).

Algumas tabelas de catálogo do *DB2* são: *SYSTABLES*, *SYSCOLUMNS*, *SY-SINDEXES*.

1. **SYSTABLES**: contém uma linha para cada tabela (base ou visão) em todo o sistema. Para cada uma dessas tabelas ela fornece informações como: nome de tabela (*NAME*), nome de usuário que o criou (*CREATOR*) e quantidade de suas colunas (*COLCOUNT*).
2. **SYSCOLUMNS**: contém uma linha para cada coluna de cada tabela em todo o sistema. Para cada uma dessas colunas, uma linha fornece, por exemplo, nome da coluna (*NAME*), nome da tabela da qual a coluna faz parte (*TBNAME*) e tipo de dados da coluna (*COLTYPE*).
3. **SYSINDEXES**: contém uma linha para cada índice no sistema. Para cada um desses índices ela fornece: nome do índice (*NAME*), nome da tabela indexada (*TBNAME*), nome de usuário que criou o índice (*CREATOR*) e assim por diante.

2.2.2 Catálogo no *MySQL*

Em um banco de dados institucional, o *INFORMATION_SCHEMA* é visto somente pelo administrador do banco de dados, pois ali estão armazenadas todas as

informações referentes aos bancos de dados criados.

A Figura 2.1 ilustra parcialmente o `INFORMATION_SCHEMA` do banco de dados `MySQL`. Ele é o banco de informações, o local que se armazena informações sobre todos os outros bancos que o servidor `MySQL` mantém [49]. Dentro do `INFORMATION_SCHEMA` existem várias tabelas somente para leitura. Elas são visões e não são tabelas de base, portanto, não existem arquivos a elas associados. O `INFORMATION_SCHEMA` fornece acesso aos metadados do banco de dados, tais como o nome de um banco de dados ou tabela, o tipo de dados de uma coluna e privilégios de acesso.

Comentário do Banco de Dados:

CHARACTER_SETS

Campo	Tipo	Nulo	Padrão	Links para	Comentários
CHARACTER_SET_NAME	varchar(64)	Não			
DEFAULT_COLLATE_NAME	varchar(64)	Não		COLLATIONS -> COLLATION_NAME	
DESCRIPTION	varchar(60)	Não			
MAXLEN	bigint(3)	Não	0		

COLLATIONS

Campo	Tipo	Nulo	Padrão	Links para	Comentários
COLLATION_NAME	varchar(64)	Não			
CHARACTER_SET_NAME	varchar(64)	Não		CHARACTER_SETS -> CHARACTER_SET_NAME	
ID	bigint(11)	Não	0		
IS_DEFAULT	varchar(3)	Não			
IS_COMPILED	varchar(3)	Não			
SORTLEN	bigint(3)	Não	0		

COLLATION_CHARACTER_SET_APPLICABILITY

Campo	Tipo	Nulo	Padrão	Links para	Comentários
COLLATION_NAME	varchar(64)	Não		COLLATIONS -> COLLATION_NAME	
CHARACTER_SET_NAME	varchar(64)	Não		CHARACTER_SETS -> CHARACTER_SET_NAME	

Figura 2.1: Dicionário de dados do banco de dados `MySQL`: `INFORMATION_SCHEMA` [55]

É possível selecionar o `INFORMATION_SCHEMA` como o banco de dados padrão com uma instrução `USE`, mas com ele só é possível ler o conteúdo das tabelas. Não é possível inserir, atualizar ou apagar dados a partir dele.

Algumas tabelas do `INFORMATION_SCHEMA` são descritas a seguir. A descrição completa do `INFORMATION_SCHEMA` do `MySQL` pode ser encontrada em [48].

1. **SCHEMATA.** Um esquema (*schema*) descreve um banco de dados, portanto a tabela `SCHEMATA` informa sobre um banco de dados;
2. **TABLES.** A tabela `TABLES` informa sobre as tabelas criadas nos bancos de dados;
3. **COLUMNS.** A tabela `COLUMNS` informa sobre colunas e tabelas de todos os bancos de dados;
4. **STATISTICS.** A tabela `STATISTICS` informa sobre tabela de índices;
5. **USER_PRIVILEGES.** A tabela `USER_PRIVILEGES` informa sobre privilégios globais, cujo acesso é obtido a partir da permissão da tabela `mysql.user`;

6. **SCHEMA_PRIVILEGES**. A tabela *SCHEMA_PRIVILEGES* informa sobre o esquema de privilégios do banco de dados obtido a partir da permissão da tabela *mysql.user*.

Um modelo conceitual do *INFORMATION_SCHEMA* do Gerenciador de banco de dados *MySQL* é apresentado na Figura 2.2. O modelo fornece informações como os índices das tabelas, os elementos de dados armazenados em cada registro, além dos graus, conectividade e atributos de um relacionamento.

2.2.3 Catálogo no *Oracle*

Os catálogos no *Oracle* [54] são informações dos objetos do banco de dados *Oracle* como pode ser visto parcialmente a seguir:

1. **ALL_TABLES**: contém todas as tabelas do banco de dados atual que são acessíveis ao usuário;
2. **ALL_TAB_COLUMNS**: contém todas as colunas no banco de dados que são acessíveis ao usuário;
3. **ALL_ARGUMENTS**: contém os argumentos de funções e procedimentos que são acessíveis ao usuário;
4. **ALL_ERRORS**: contém descrições de erros sobre todos os objetos armazenados - visões, procedimentos, funções, pacotes e corpos de pacotes - que são acessíveis ao usuário;
5. **ALL_OBJECT_SIZE**: incluído para fins de compatibilidade a versões anteriores do *Oracle 5*;
6. **ALL PROCEDURES**: (a partir do *Oracle 9* em diante) contém todas as funções e os procedimentos (juntamente com as propriedades associadas) que são acessíveis ao usuário atual;
7. **ALL_SOURCE**: descreve o texto (isto é, o PL/SQL), fonte de objetos armazenados acessíveis ao usuário.

Alguns itens do catálogo do *PostgreSQL* são descritos a seguir. O Catálogo completo do SGBD *PostgreSQL* pode ser encontrado em [57].

1. **pg_database**: contém bancos de dados que fazem parte do agrupamento de bancos de dados;
2. **pg_type**: contém tipos de dado;
3. **pg_depend**: contém dependências entre objetos dos bancos de dados;
4. **pg_description**: contém descrições ou comentários sobre os objetos dos bancos de dados;
5. **pg_group**: contém grupos de usuários dos bancos de dados;
6. **pg_index**: contém informações adicionais sobre índices.

Nos exemplos de catálogos dos bancos de dados apresentados, observa-se que cada banco armazena os metadados do esquema de forma bastante específica, o que torna trabalhosa a extração dos metadados.

A próxima seção é destinada ao Protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). Este protocolo tem sido utilizado para a interoperabilidade entre os repositórios digitais com o objetivo de promover o acesso simultâneo aos dados nele contidos.

2.3 Open Archives Initiative

A *Open Archives Initiative* (OAI) foi criado para facilitar a disseminação eficiente de conteúdo, com o objetivo de desenvolver e promover padrões de interoperabilidade entre repositórios digitais. Com essa filosofia, a OAI criou o protocolo OAI-PMH, um meio simples de se compartilhar metadados entre servidores distribuídos [5].

Ao invés de fazer buscas em diversas bases de dados no momento da consulta pelo usuário, o protocolo OAI baseia-se na busca automática de metadados (*harvesting*) de recursos que adotam o protocolo e a implementação posterior de uma base de dados, que armazena os metadados coletados. A vantagem é que, diferentemente da busca federada, o acordo entre fornecedores de recursos digitais e fornecedores de serviços se baseia num conjunto reduzido de dados, permitindo serviços básicos com um mínimo de esforço (TAMMARO e SALARELLI, 2008, p. 247) [64].

O movimento OAI tem suas raízes no esforço de ampliar o acesso a repositórios de artigos científicos como um meio de aumentar a disponibilidade da comunicação científica. Entretanto, os principais padrões e ferramentas tecnológicas desenvolvidos são independentes do tipo de conteúdo oferecido. A OAI não se restringe apenas a repositórios científicos, mas também a repositórios de músicas, *Uniform Resource Locator* (urls) e vídeos [5].

Dentre as características da *Open Archives Initiative*, que especificam sua usabilidade e disponibilidade estão: autoarquivamento, interatividade, interdisciplinaridade e interoperabilidade [41].

O autoarquivamento é a possibilidade de o autor ter autonomia para editar e arquivar seus textos. Essa liberdade permite em alguns casos ratificações e atualizações da obra, valorizando o conteúdo informacional constantemente. A interatividade é uma característica que diz respeito à interação entre a comunidade acadêmica, possibilitando troca de informações, críticas, comentários e sugestões do texto disponível nesse sistema. Essa forma de comunicação garante legitimidade e aponta o prestígio e a aceitação das informações contidas no texto. Isso acontece por meio de convergências e confrontos de ideias, discussões, críticas e consensos entre a comunidade participante.

A interdisciplinaridade refere-se à diversidade de áreas do conhecimento em um repositório digital e por fim a interoperabilidade objetiva configurar todos os arquivos, padronizando-os globalmente para que o acesso a eles seja ilimitado, sem restrições de plataformas [41].

Essas características apontam pontos positivos, apesar da utilização do modelo de comunicação *Open Archives* apresentar questões em aberto na comunidade científica. Essas questões estão relacionadas à absorção dos custos de criação e às dificuldades de aceitação como item de produtividade, conforme apresentado por Machado (2006, p. 42-43) [41].

A absorção dos custos de criação dos *Open Archives* desperta a questão de como as instituições interessadas poderão custear os gastos com material tecnológico, humano e com infraestrutura, além da manutenção, divulgação e muitas vezes incerteza da continuidade das iniciativas que oferecem *software* livre para implementação dos *Open Archives*.

Em relação aos direitos autorais, a *Open Archives Initiative* [27] discorre que "o gerenciamento da propriedade intelectual dos conteúdos dos documentos digitais são de responsabilidade do Provedor de Dados ou de Serviço que adota o protocolo OAI-PMH."

Este fato levanta indagações sobre a confiabilidade nos conteúdos dos trabalhos disponibilizados nos *Open Archives*. Apesar de suas características serem positivas, a OAI é uma iniciativa que foi criada em 1999 e, por ser considerada nova no ambiente científico, sua existência e suas atividades levantam interrogações acerca de sua utilização como alternativa de comunicação e apoio de escrita para disseminação e publicação da produção científica nacional e internacional.

2.4 O Protocolo OAI-PMH

As definições e as descrições do Protocolo OAI-PMH apresentadas neste trabalho foram baseadas no sítio da Incubadora da FAPESP [22] e no sítio oficial que trata do Protocolo OAI-PMH versão 2.0 [53].

O *Open Archives Initiative Protocol* para coleta de metadados (OAI-PMH) é um mecanismo de pouco obstáculo para a interoperabilidade de repositórios e para transferência de dados entre repositórios de bibliotecas digitais. É uma interface que um servidor de rede pode empregar para que os metadados de objetos residentes no servidor estejam disponíveis para aplicações externas que desejam coletar esses dados. Essa interface tem duas propriedades: interoperabilidade e extensibilidade.

No protocolo existem duas classes de participantes: "provedores de dados" e "provedores de serviços". Os provedores de dados são repositórios que expõem metadados estruturados por meio de OAI-PMH, enquanto os provedores de serviços fazem solicitações de serviços OAI-PMH para a colheita de metadados.

O protocolo utiliza um ferramental de transações *Hypertext Transfer Protocol* (HTTP) simples, baseadas em pergunta-resposta (*request-response*). A resposta deve ser no formato XML. Os *harvesters* são programas que utilizam a interface oferecida pelo protocolo OAI-PMH para realizar a coleta de metadados. Um *harvester* pode selecionar os metadados que deseja coletar de forma a obter somente os novos ou os que foram modificados desde a última interação de coleta com o repositório e também restringir os metadados que se deseja coletar ao indicar qual o conjunto de dados de interesse (alguns servidores separam os metadados por assunto ou por ano de publicação). O processo descrito aparece ilustrado na Figura 2.3.

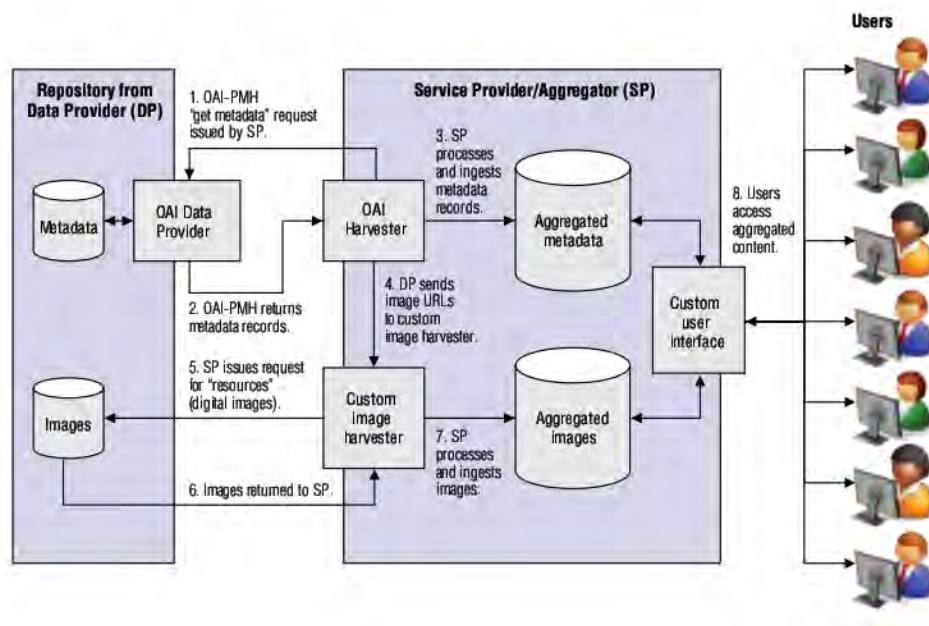


Figura 2.3: Metadata harvesting model [28]

As requisições dos provedores de serviços são realizadas em formato HTTP, através de uma URL básica do repositório mantido pelo provedor de dados. Além da URL básica para coleta, é preciso identificar o que será coletado e como a mesma será realizada (i.e. identificar o tipo de *harvesting*). Para isto, o OAI define seis verbos que especificam detalhes da coleta dos repositórios e alguns argumentos, a fim de refinar o *harvester* [13]. Os verbos de requisição são comandos transmitidos aos repositórios do protocolo OAI-PMH e que por sua vez retornam as respostas codificadas em XML. São eles [23]:

- *Identify*: apresenta as principais informações do repositório: nome, identificador, *e-mail* do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc.
- *ListMetadataFormats*: lista os formatos de metadados implementados pelo sistema do repositório; é obrigatória a implementação de pelo menos o formato *Dublin Core*.
- *ListSets*: expõe a árvore de assuntos que classifica os documentos no repositório ou outro conjunto de classificação.
- *ListIdentifiers*: lista todos os identificadores de registros do repositório.
- *ListRecords*: lista todos os registros do repositório.
- *GetRecord*: lista o registro correspondente, dado um identificador.

A tabela a seguir destaca os verbos e os argumentos utilizados para refinar a coleta de dados.

Tabela 2.2: Os verbos e seus argumentos

[13]

Verbo	Descrição	Argumentos
GetRecord	Recupera os metadados de um item individual de um repositório.	identifier . Obrigatório. Deve ser especificado o identificador único do item de um repositório. metadataPrefix . Obrigatório. Especifica o padrão de metadados adotado que deve estar especificado no provedor de dados.
Identify	É usado para coletar informações sobre um repositório.	Não há argumentos.
ListRecords	Este verbo recupera os metadados de um repositório.	from . Opcional. Os dados coletados devem ser criados ou alterados a partir da data especificada por este argumento. until . Opcional. Os dados devem ser criados ou alterados até a data especificada pelo argumento. metadataPrefix . Explicado em GetRecord. set . Opcional. Especifica um conjunto, para o <i>harvester</i> poder refinar a coleta. resumptionToken . Exclusivo. Argumento necessário quando os provedores utilizam o controle de fluxo na coleta dos metadados.
ListIdentifiers	É uma abreviação do <i>ListRecords</i> , que retorna apenas o <i>header</i> de um repositório.	from . until . metadataPrefix . set . resumptionToken .
ListMetadataFormats	Retorna os padrões de metadados utilizados em um repositório.	identifier . Opcional (apenas neste verbo). Retorna o padrão de metadados utilizado em um item específico.
ListSets	É utilizado para retornar a estrutura de um repositório, listando todos os conjuntos que compõe os metadados.	resumptionToken .

2.4.1 Objetivo do Protocolo OAI-PMH

O Protocolo *Open Archives Initiative - Protocol for Metadata Harvesting* (OAI-PMH) foi criado com o objetivo de facilitar a disseminação eficiente de conteúdos entre repositórios digitais, tornando-se um padrão para o desenvolvimento e a ampliação das atividades de pesquisa científica.

A principal finalidade do protocolo OAI-PMH é facilitar o compartilhamento de metadados, ou seja, a coleta e a agregação desses dados para criação de novos serviços. O exemplo mais comum é a busca federada.

O termo "busca federada" possui vários sinônimos: metabusca, busca distribuída, busca paralela, portal de busca, *broadcast search*, *cross-database search*. Ao adotar a expressão mais frequentemente utilizada - busca federada - entende-se que essa busca consiste na transmissão de uma pergunta com a sintaxe apropriada a um grupo de bases de dados dispersas, agrupando os resultados coletados das bases de dados e apresentando-os em um formato sucinto e unificado, com duplicação mínima, permitindo que o usuário

da biblioteca classifique o resultado agrupado ajustando-o por meio de vários critérios (JACSO, 2004, p. 17) apud Ferreira [25].

Wadham (2004, p. 20) apud Ferreira [25] define a busca federada como uma técnica pela qual as fontes de informação múltiplas podem ser pesquisadas ao mesmo tempo, incluindo catálogos de bibliotecas, bases de dados comerciais, mecanismos de busca *web* e outros metadados ou bases de dados digitais. O mecanismo de busca federada, usando um protocolo de busca, tal como o Z39.50 por exemplo, recolhe a informação a partir de várias fontes e fornece um resultado de busca integrado ao usuário (Figura 2.4A). Na coleta de metadados/*harvesting* existe uma cópia local de vários repositórios em que os metadados são coletados e salvos e a consulta a esses metadados é realizada neste repositório local (Figura 2.4B). O modelo de busca federada e de coleta de metadados/*harvesting* está representada na Figura 2.4.

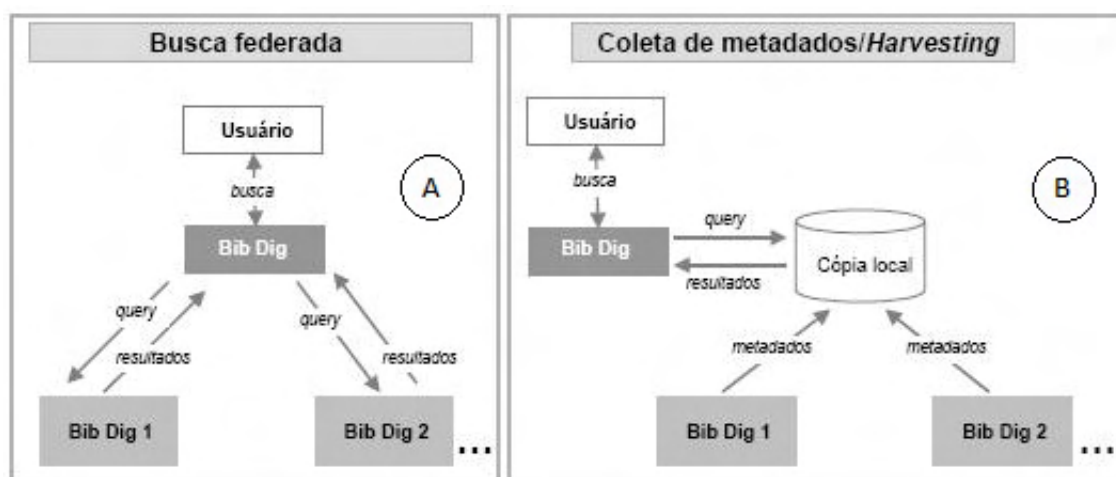


Figura 2.4: Modelo de busca federada e de coleta de metadados/*harvesting* (SULEMAN, 2002, p. 15) apud Ferreira [25]

A próxima subseção apresenta as etapas para uma coleta eficiente dos metadados.

2.4.2 Etapas para uma Coleta Eficiente de Metadados

A FAPESP [24] lista alguns passos a serem seguidos para uma coleta eficiente de metadados:

1. Selecionar os repositórios dos quais se deseja coletar os metadados e descobrir as respectivas urls bases de coleta (ou seja, as urls a partir das quais é possível aplicar os verbos de requisição);
2. Escolher uma ferramenta apropriada de *harvesting*;
3. Iniciar a coleta, fornecendo uma lista com as urls bases dos repositórios;

4. Normalizar os metadados coletados de forma a se tornarem úteis para implementação do serviço pretendido.

Apesar dos passos listados possibilitarem uma coleta eficiente de metadados, algumas dificuldades são apontadas em cada uma dessas etapas. A seguir são detalhadas essas dificuldades [24]:

1. **Seleção**

Decidir quais os repositórios cujos metadados participarão do seu serviço é uma tarefa manual e exaustiva. A partir da resposta do verbo *Identify*, é possível descobrir algumas informações sobre a natureza dos dados contidos no repositório, porém, frequentemente, esses dados são insuficientes para se saber se é de interesse a coleta desses metadados. Para tal decisão é necessário navegar nos recursos do repositório e também verificar a qualidade da implementação dos metadados. Desenvolver ferramentas computacionais que automatizem essa etapa é uma proposta bastante complexa, principalmente porque o tema e a qualidade do repositório são critérios subjetivos e sujeitos a debates. Os diversos repositórios não só implementam o padrão *Dublin Core* de forma diferente, mas também apresentam diferentes padrões de qualidade.

2. **Ferramentas para *harvesting***

Um serviço bastante útil para testar se o repositório de interesse está respondendo adequadamente aos verbos OAI é o *Repository Explorer*.

O *Repository Explorer* [61] é uma ferramenta *web* inicialmente desenvolvida pela *Digital Library Research Laboratory* (DLRL) da *Virginia Tech*, que possibilita a verificação interativa da validade da implementação do protocolo OAI-PMH para um determinado repositório. A ferramenta é mantida pela *Advanced Information Management Laboratory* (AIM) da University of Cape Town.

Para testar o serviço, basta fornecer a ele a url base do repositório e preencher alguns campos relativos aos verbos de requisição. Depois de testados, é necessário fazer a coleta dos metadados. Para tal fase, é imprescindível escolher uma ferramenta adequada às suas necessidades. No portal de ferramentas da OAI, pode-se encontrar alguns *harvesters* de licença livre. Suas funcionalidades e desempenho variam, por isso devem ser bem testados de acordo com seus requisitos. Uma decisão mal feita nessa etapa pode atrapalhar bastante a implementação do serviço pretendido.

3. **Coleta de metadados**

Os serviços de coleta de metadados podem ser interrompidos em função dos seguintes fatores:

- (a) **Mudança de URL base do repositório:** para coletar via protocolo OAI é preciso fornecer ao programa *harvester* uma URL a partir da qual ele aplicará os verbos de requisição; porém, às vezes a url do repositório muda e é difícil descobrir a nova URL;
- (b) **Servidor "fora do ar":** quando se utiliza *software* livre como *EPrints* ou *DSpace*, os repositórios institucionais são baratos e fáceis de se montar. Por isso, algumas vezes são implementados por instituições que não possuem políticas para manutenção e sustentação do serviço;
- (c) **Falhas na implementação da interface OAI:** alguns repositórios não respondem corretamente aos seis verbos do protocolo;
- (d) **Interrupção na coleta por sobrecarga do servidor:** por receberem muitas requisições de *harvesters*, alguns servidores ficam sobrecarregados e interrompem a transferência dos metadados antes de ser finalizada. Esses pequenos empecilhos tornam-se mais incômodos à medida que cresce o número de repositórios dos quais se fazem a coleta de metadados, tornando a etapa de *harvesting* um tanto trabalhosa e exaustiva.

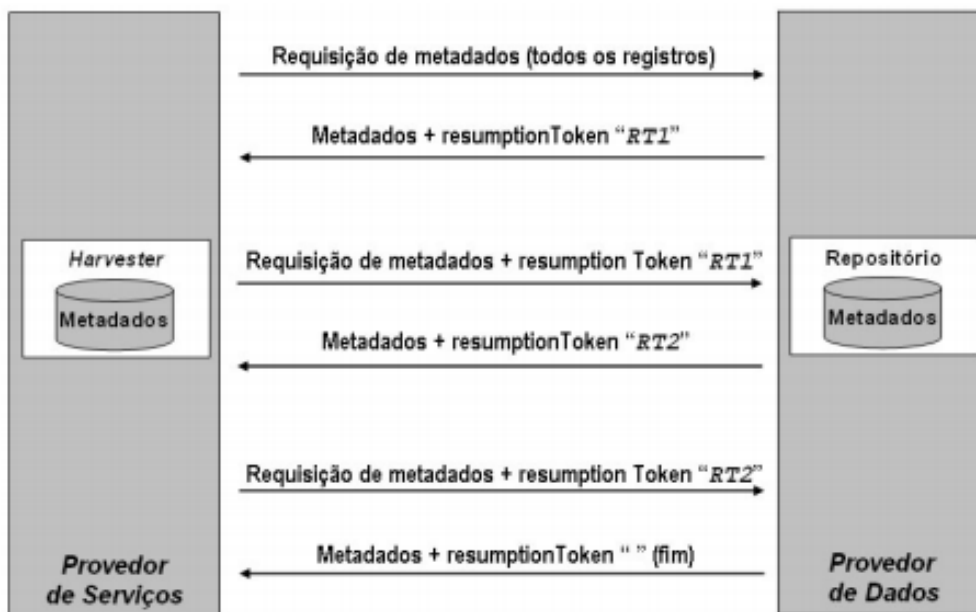


Figura 2.5: Funcionamento básico do controle de fluxo do OAI-PMH [8]

4. Heterogeneidade nos metadados e necessidade de normalização

O padrão *Dublin Core* pode ser implementado com vários tipos de dados, porém, para ser útil, os metadados precisam ser adaptados de forma a se adequarem ao serviço. Como os metadados nesse padrão não pretendem descrever apenas *eprints*, a definição de alguns campos apresentada no *site* oficial do *Dublin Core* é um tanto

vaga. Por exemplo, o campo *relation* - de acordo com a definição dada no *Dublin Core* - é uma referência a um recurso relacionado.

Na realidade, o protocolo OAI não implementa exatamente o *Dublin Core*, mas sim o *oai_dc*. Trata-se de um subconjunto de campos do *Dublin Core*. A principal diferença entre o *Dublin Core* e o *oai_dc* é que neste a implementação de alguns campos se torna obrigatória. São eles: *title, creator, subject, description, publisher, date, type, identifier, relation, language*.

Existem algumas recomendações de implementação desses campos para *eprints*, feitas por Andy Powell, Michael Day e Peter Cliff (2004). Por exemplo, recomenda-se colocar a URL para o texto completo no campo *relation*. As recomendações são bastante boas do ponto de vista das informações necessárias para que tanto um robô quanto um ser humano sejam capazes de extrair dados importantes para avaliação do documento. Porém, elas são muito pouco seguidas mesmo pelos maiores e mais importantes repositórios de artigos científicos de Acesso Aberto.

Verifica-se que em repositórios famosos como *arXiv, Pubmed Central, CiteSeer, Cogprints*, o repositório *ECS EPrints* entre outros, os campos *title* e *description* são frequentemente bem implementados, seguindo as recomendações de Powell, Day e Cliff. Eles contém respectivamente o título e o resumo do artigo. No entanto, há muita confusão no campo *date*, que pode se referir à data de publicação na revista, à disponibilização do artigo no repositório digital ou até mesmo à data de criação do documento.

Além desses quatro campos importantes mencionados no parágrafo anterior para a criação de uma biblioteca temática, os campos *subject* e *set* (assunto e conjunto) assumem um papel especial para ajudar na seleção e separação dos artigos por assunto. Porém, surgem muitas dificuldades devido à ausência de padrões compartilhados entre os repositórios.

Pode-se observar formas de utilização do campo *set* bastante diferentes ao observar o repositório institucional de *Southampton*, o *Cogprints* e o *DSpace* do MIT. O repositório de *Southampton*, por exemplo, implementa o difundido padrão de classificação *Library of Congress*. O *Cogprints* utiliza um esquema especializado nos ramos das suas disciplinas (Ciências Cognitivas e Ciência da Computação). Já o *DSpace*, hospedado pelo MIT, utiliza nos campos *subject* as palavras-chave que o autor escolheu para o documento. Somente nesses três exemplos tem-se três esquemas completamente diferentes.

Nas suas últimas versões, o *software EPrints* tem incorporado o padrão de classificação *Library of Congress*. Trata-se de um dos melhores e mais famosos, porém, para muitas áreas especializadas da Ciência essa árvore não é suficientemente ramificada, já que abrange apenas os assuntos mais gerais.

Essas deficiências podem ser diminuídas por meio de ferramentas computacionais sofisticadas. Uma possível solução seria, a partir do texto completo, encontrar automaticamente as palavras-chave do artigo e utilizá-las para inferir o assunto. Ainda não existe uma ferramenta desse tipo entre as de apoio da OAI.

Ao longo do texto verificou-se que o protocolo OAI-PMH desenvolve e promove padrões de interoperabilidade entre repositórios digitais. Existem várias definições para eles que são apresentadas na seção a seguir.

2.5 Repositórios Digitais

O protocolo OAI-PMH tem sido utilizado para a interoperabilidade entre os repositórios digitais com o objetivo de promover o acesso simultâneo aos dados neles contidos de forma a maximizar as buscas e reduzir o tempo de resposta [16].

Para um conhecimento mais aprofundado do que seriam os repositórios digitais, a seguir são apresentadas algumas definições:

Segundo Leite (apud MASSON, 2008, p. 137) [45],

repositórios digitais são uma consequência da aplicação da Internet e de tecnologias emergentes de informação e comunicação no contexto da comunicação científica, somada ao movimento mundial de acesso livre à informação científica, fundamentado no modelo de arquivos abertos, levando ao surgimento de modelos alternativos que provocaram mudanças paradigmáticas no processo de comunicação científica.

Já para Viana et al. (apud MASSON, 2008, p. 137) [45]:

Um repositório digital é uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos de tempo e prover o acesso apropriado. Essa estratégia foi possibilitada pela queda nos preços no armazenamento, pelo uso dos padrões como o protocolo de coleta de metadados da Iniciativa dos Arquivos Abertos (OAI-PMH), e pelos avanços no desenvolvimento dos padrões de metadados que dão suporte ao modelo de comunicação dos arquivos abertos.

Para Weitzel apud Masson (2008, p. 137) [45],

o repositório de *E-prints* [é] um tipo de repositório digital, que pode ser temático ou institucional, mas que utiliza obrigatoriamente o *software E-prints* desenvolvido pela University of Southampton, cujos textos que reúne já foram publicados ou não, e estão sujeitos à crítica por meio de comentários de outros pesquisadores.

Ainda segundo Weitzel apud Masson (2008, p. 137) [45]:

Os provedores de dados são os repositórios digitais, propriamente ditos, incluindo os repositórios de *e-prints*. Também são as publicações *on-line* individuais, isto é, periódicos científicos, teses e dissertações, anais de eventos etc. contendo, preferencialmente, os textos completos dos trabalhos já publicados como os seus dados descritivos (metadados).

Para Reese et al. (2008, p. 1) [35],

Um repositório digital é simplesmente uma coleção de fontes digitais [35].

Para entender o funcionamento dos repositórios digitais, a Figura 2.6 representa de forma simples o fluxo de execução de um deles.

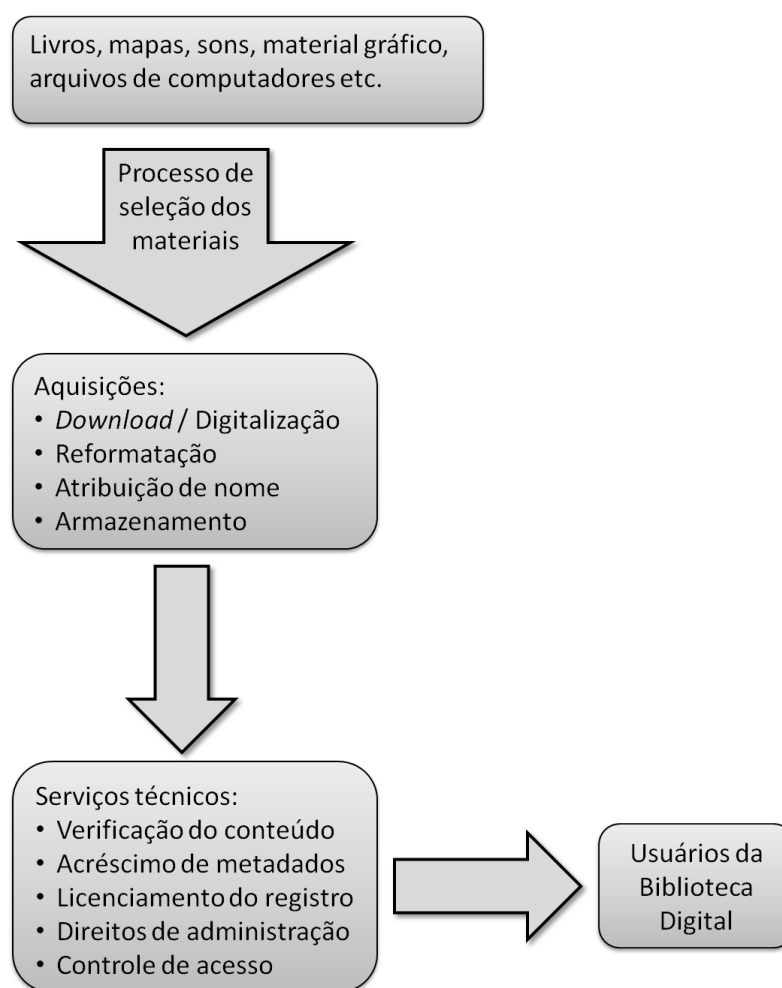


Figura 2.6: Fluxo de execução de um repositório digital. Adaptado de [35]

As bibliotecas dependem de uma variedade de procedimentos para incorporar materiais como livros, jornais e documentos governamentais em suas coleções. O processo de desenvolvimento da coleção determina qual material será incluído ou removido da coleção [35]. O fluxo de um repositório digital apresentado na Figura 2.6 define o

processo físico e conceitual que ocorre quando as bibliotecas adquirem novos materiais, organizam ou administram as coleções e promovem o seu acesso.

Os novos materiais que as bibliotecas recebem ou adquirem passam por um processo de seleção acompanhado por especialistas, que os identificam e os catalogam. Nesse processo de seleção, os materiais podem passar por diversos profissionais que são designados para realizar tarefas específicas, pois um fluxo de trabalho elaborado especifica as pessoas e as habilidades necessárias para executar as tarefas do fluxo. Para um fluxo contínuo de seleção de novos materiais seriam necessários especialistas para cada atividade - seleção de materiais feita por meio de *download* ou digitalização, reformatação do material, atribuição e identificação do material e depois seu armazenamento.

Ao passar pelo processo de seleção, o próximo passo são os serviços técnicos que já exigem outras habilidades dos especialistas - verificação do conteúdo, inclusão de metadados, licenciamento do registro, administração de direitos e controle de acesso. Somente após passar por essas etapas o material é disponibilizado para os usuários da biblioteca.

Ao criar um repositório digital de banco de dados tornar-se-á possível a interoperabilidade entre os repositórios, além de possibilitar a disseminação de dados e a expansão de consultas a banco de dados na *web*.

2.6 Dublin Core

O padrão metadados *Dublin Core* é um conjunto eficaz para descrever uma ampla gama de recursos de rede. Esse padrão inclui dois níveis: Simple e Qualificado. O nível Simple inclui quinze elementos discutidos na Seção 2.6.2, o nível Qualificado, três elementos adicionais (Audiência, Proveniência e Detentor de Direitos), assim como um grupo de refinamentos de elementos (também chamados qualificadores), que refinam a semântica dos elementos de maneira que podem ser úteis para descoberta de recursos [19].

A semântica do *Dublin Core* tem sido estabelecida por um acordo internacional de grupo transdisciplinar de profissionais de biblioteconomia, ciência da computação, codificação de texto, comunidade de museus e outras áreas afins de estudos e práticas.

Outra maneira de olhar para *Dublin Core* é como uma "pequena linguagem para fazer declarações de uma classe particular sobre recursos". Nessa linguagem há duas classes de termos - elementos (substantivos) e qualificadores (adjetivos) - que podem ser organizadas em um padrão simples de instrução. Os recursos próprios são os sujeitos implicados nesse idioma. (Uma discussão adicional da Gramática de *Dublin Core* encontra-se em "Princípios gramaticais de *Dublin Core Metadata Initiative* (DCMI)"). No

mundo diverso da *internet*, *Dublin Core* pode ser visto como um "pidgin"¹ metadados para turistas digitais": facilmente compreendida, mas não necessariamente à altura da tarefa de expressar relações complexas ou conceitos.

Cada elemento do *Dublin Core* é opcional e pode ser repetido. A maioria dos elementos tem também um conjunto limitado de qualificadores ou refinamentos, atributos que podem ser utilizados para refinar ainda mais (e não estender) o significado do elemento. O *Dublin Core Metadata Initiative* (DCMI) estabeleceu maneiras padronizadas para refinar os elementos e encorajar o uso de esquemas de codificação e vocabulário.

2.6.1 Objetivos do *Dublin Core*

O *Dublin Core* tem como objetivos:

1. **Simplicidade de criação e manutenção:** o conjunto de elementos *Dublin Core* foi mantido tão pequeno e simples quanto possível para permitir que um não-especialista crie simples registros descritivos para os recursos de informação fáceis e de custos acessíveis, enquanto prevê a recuperação efetiva desses recursos no ambiente de rede.
2. **Semântica comumente entendida:** descoberta de informação por meio da *internet* é dificultada pelas diferenças de terminologia e práticas descritivas de um campo de conhecimento para o outro. O *Dublin Core* pode ajudar o turista "digital", não um pesquisador especializado, a encontrar o seu caminho por meio do amparo a um conjunto comum de elementos, a semântica do que é universalmente compreendido e apoiado. Por exemplo, cientistas preocupados com a localização de artigos de um determinado autor, e os estudiosos interessados em obras de arte de um determinado artista podem concordar com a importância de um elemento "criador". Essa convergência de um conjunto de elemento comum, um pouco mais genérico, aumenta a visibilidade e acessibilidade de todos os recursos, não somente dentro de uma determinada disciplina.
3. **Âmbito internacional:** o conjunto de elementos *Dublin Core* foi desenvolvido originalmente em inglês, mas outras versões estão sendo criadas em muitas outras línguas, incluindo o finlandês, norueguês, tailandês, japonês, francês, português, alemão, grego, indonésio e espanhol. O grupo de interesse especial de localização e internacionalização do DCMI coordena esforços para vincular essas versões em

¹Pidgin: uma língua que é uma mistura de outras duas línguas e é utilizada entre pessoas que não falam bem uma língua entre eles.

um registro distribuído.

4. **Extensibilidade:** apesar de equilibrar as necessidades para a simplicidade na descrição de recursos digitais com a necessidade de recuperação precisa, os desenvolvedores *Dublin Core* têm reconhecido a importância de fornecer um mecanismo para estender o conjunto de elementos DC para a necessidade de descoberta de fonte adicional. Espera-se que outras comunidades de peritos em metadados criem e administrem conjuntos de metadados adicionais, especializados para atender às necessidades de suas comunidades. Elementos de metadados a partir desses conjuntos poderiam ser usados em conjunto com os metadados *Dublin Core* para satisfazer a necessidade de interoperabilidade. O DCMI *Usage Board* está presentemente trabalhando num modelo para realizar isso no contexto de perfis de "aplicação".

2.6.2 Elementos do *Dublin Core*

As descrições dos elementos de metadados são normalizadas. Essa normalização ajuda a melhorar a consistência com outras comunidades de metadados, permitindo ainda uma melhor clarificação, definição do alcance e consistência interna nas definições dos elementos de metadados *Dublin Core* [19].

Cada elemento *Dublin Core* é definido usando um conjunto de dez atributos da norma ISO/IEC 11179 [33] para a descrição de elementos de dados. Esses consistem em:

1. **Nome:** a etiqueta atribuída ao elemento de dado;
2. **Identificador:** o identificador único atribuído ao elemento de dado;
3. **Versão:** a versão do elemento de dado;
4. **Autoridade:** a entidade autorizada a registrar o elemento de dado;
5. **Língua:** a linguagem na qual o elemento de dado é definido;
6. **Definição:** uma afirmação que representa claramente o conceito e a natureza do elemento de dado;
7. **Obrigaçãõ:** indica se o elemento de dado é sempre obrigatório ou não (contém um valor);
8. **Tipo dos dados:** indica o tipo de dados que pode ser representado no valor do elemento de dado;
9. **Máxima ocorrência:** indica qualquer limite à repetição do elemento de dado;
10. **Comentário:** uma nota relativa à aplicação do elemento de dado.

Seis dos referidos dez atributos são comuns a todos os elementos *Dublin Core*:

Exemplo:

1. **Versão:** 1.1;

2. **Autoridade de Registro:** *Dublin Core Metadata Initiative*;
3. **Língua:** pt ("en" para a versão original);
4. **Obrigaç o:** opcional;
5. **Tipo dos Dados:** cadeia de caracteres;
6. **M xima Ocorr ncia:** ilimitada.

O atributo "Defini o" captura o conceito sem ntico enquanto os atributos "Tipo dos Dados" e "Coment rio" capturam a representa o dos dados. Por sua vez, o atributo "Identificador" dever  ser entendido n o como um nome, mas apenas como uma etiqueta ou c digo  nico identificador do elemento, o qual ser  sempre representado pela palavra em l ngua inglesa, equivalente ao valor do atributo "Nome" (etiqueta que dever  ser identificada pelas aplica es e sistemas que interpretem e processem registros de metadados *Dublin Core*. Por isso, h  a necessidade de se manter numa forma  nica e compat vel para qualquer aplica o, independentemente da l ngua do contexto).

Na Tabela 2.3 s o descritos os quinze elementos do *Dublin Core Simple*s:

Tabela 2.3: *Elementos do Dublin Core [26]*

Elemento	Descri�o
Contributor	Pessoas que contribuíram para o trabalho
Coverage	�reas de abrang�ncia do trabalho
Creator	Autores do trabalho
Date	Data de disponibiliza�o do trabalho
Description	Descri�o do trabalho
Format	Formato em que � disponibilizado o trabalho
Identifier	Identificador �nico para o trabalho
Language	Idioma do trabalho publicado
Publisher	Informa�es sobre o respons�vel pela publica�o do trabalho
Relation	Relacionamento com outros trabalhos, se houver
Rights	Descri�o da propriedade intelectual ou direitos autorais
Source	Outras fontes das quais o trabalho se deriva, se houverem
Subject	Assunto do qual trata o trabalho
Title	T�tulo de um trabalho
Type	Tipo do objeto

Cada elemento *Dublin Core* refere-se ao recurso a ser descrito. Um recurso   definido como "tudo o que tiver uma identidade". Para o presente efeito, um recurso ser  tipicamente um servi o ou um recurso de informa o, mas o conceito poder  ser aplicado de forma mais ampla. Uma descri o detalhada dos elementos do *Dublin Core* encontra-se em [19].

Um exemplo de registro *Dublin Core* referente   Tabela 1.1   codificado como parte de uma p gina *web* usando o tag <META>:

Código 2.1 – Código HTML -Codificação do *Dublin Core* como parte de uma página web usando o tag <META>

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
2 <HTML>
3 <HEAD>
4 <TITLE>Exemplo de um simples registro de metadados</TITLE>
5 <META NAME="DC.Title"      CONTENT="Banco de Dados PostgreSQL">
6 <META NAME="DC.Creator"    CONTENT="Elisabete Tomomi Kowata">
7 <META NAME="DC.Publisher"  CONTENT="Universidade Federal de Goiás">
8 <META NAME="DC.Type"      CONTENT="text">
9 <META NAME="DC.Date"      CONTENT="09/06/2010">
10 <META NAME="DC.Format"    CONTENT="text/html">
11 <META NAME="DC.Identifier" CONTENT="http://www.ufg.br/sghbd/
12 meta/mainmenu.html">
13 <META NAME="DC.Relation"   CONTENT="Gerenciador de conteúdo
14 Joomla - versão 1.6">
15 </HEAD>
16 <BODY><PRE>
17 O conteúdo disponibilizado é gerenciado por Joomla - versão 1.6
18 e utiliza o banco de dados PostgreSQL. O banco de dados contém
19 dados relacionados à administração acadêmica dos cursos ofere-
20 cidos pelo Instituto de Informática da UFG.
21 </PRE></BODY>
22 </HTML>
```

Além dos conceitos apresentados, conhecer as ferramentas que automatizem o processo de extração de metadados é importante para destacar as características que se deseja imprimir no protótipo a ser desenvolvido. Nesse sentido, algumas ferramentas que apoiam a extração de metadados também foram investigadas e são descritas na próxima seção.

O capítulo seguinte apresenta uma proposta de um método para descrever os conteúdos de bancos de dados relacionais.

Extração de Metadados de Bancos de Dados Relacionais

Se existir uma opção entre mudar e provar que alguém não tem de mudar, a maioria das pessoas ficaria ocupada na decisão. Milton Friedman, economista, 1912-?

A publicação e compartilhamento de coleções digitais são utilizadas de forma mais ampla por bibliotecas digitais e sítios da *internet*. Porém, uma vasta quantidade de informação disponível no mundo permanece armazenada em bancos de dados, principalmente relacionais. Quando um usuário submete uma consulta na *internet*, utilizando algum programa de busca, o mecanismo de execução da pesquisa não alcança aquelas informações armazenadas em bancos de dados, uma vez que esses ou são restritos às organizações, que por algum motivo não querem (ou não podem) dispor de seus dados, ou inexistente um mecanismo que permita a divulgação dessas fontes de informação. Assim, a publicação e compartilhamento de metadados de bancos de dados são uma vertente que ainda não foi explorada, o que torna o trabalho relevante no campo dos sistemas de informação.

Com a finalidade de incluir as fontes de informação estruturadas como bancos de dados relacionais, este capítulo propõe um mecanismo que permite a inclusão desses bancos como fontes de informação em pesquisas pela *internet*. O mecanismo proposto permite a consulta a bibliotecas digitais e a bancos de dados de forma homogênea por meio da definição de um conjunto comum de metadados, utilizados para exportar (disponibilizar) essas fontes de informação.

Inicialmente é feito um estudo dos elementos do padrão *Dublin Core* utilizados para representar metadados para bibliotecas digitais, visando elucidar os elementos utilizados e o significado representado por cada um deles. Em seguida, são apresentados os elementos para estabelecer a correspondência com o padrão escolhido, de forma a contemplar os campos relevantes para a descrição dos metadados de bancos de dados. O conjunto de elementos do padrão *Dublin Core* escolhido para descrever os conteúdos dos bancos de dados é, então, estruturado como uma extensão do catálogo de um sistema

gerenciador de banco de dados relacionais. No mecanismo proposto, esses metadados são extraídos da extensão do catálogo e exportados utilizando o protocolo OAI-PMH, permitindo, dessa forma, a identificação e a localização dos bancos de dados disponibilizados na rede.

Ao final é apresentado um exemplo de descrição do conteúdo de um banco de dados fictício.

3.1 Definição dos Elementos do *Dublin Core*

O catálogo de um sistema gerenciador de banco de dados contém tabelas que armazenam informações como: esquemas de bancos de dados criados no sistema, descrição das tabelas componentes de cada esquema, definição das colunas que compõem cada tabela, definição de restrições, usuários de cada banco de dados com as respectivas permissões de acesso etc. Entretanto, um breve estudo dos catálogos dos principais gerenciadores de banco de dados disponíveis atualmente (Capítulo 2) demonstra que os metadados utilizados para descrever um recurso de informação em repositórios digitais não estão disponíveis naqueles sistemas.

Assim, a fim de promover interoperabilidade de bancos de dados com outras fontes de informação, tais quais as bibliotecas digitais, é necessário definir um conjunto comum de metadados para descrever o conteúdo de um banco de dados e disponibilizá-lo (exportá-lo) por meio de um protocolo também comum.

Antes de apresentar o mecanismo proposto neste trabalho, é necessário promover um estudo sobre fontes de informação que utilizam o padrão *Dublin Core*, em especial as bibliotecas digitais, para determinar os elementos mais utilizados e seus respectivos significados.

Este estudo é composto de três etapas para a criação da extensão do catálogo do banco de dados:

1. Levantar as bibliotecas digitais, que utilizam o padrão *Dublin Core* para descrição dos metadados;
2. Determinar os elementos *Dublin Core* mais utilizados por essas bibliotecas;
3. Determinar o significado dos elementos do *Dublin Core* adotados pelas bibliotecas digitais.

As próximas subseções descrevem cada uma dessas etapas.

3.1.1 Elementos *Dublin Core* em Bibliotecas Digitais

Com o propósito de identificar os elementos *Dublin Core* mais utilizados para definir metadados em sistemas de bibliotecas digitais e seus respectivos significados, foi

realizado um levantamento com 41 bibliotecas digitais de universidades e de instituições responsáveis pela difusão de conhecimentos. O principal critério utilizado para a escolha desses repositórios está relacionado ao uso do padrão *Dublin Core* para especificar metadados utilizados para a descrição de documentos. Para se chegar a lista de bibliotecas digitais foram seguidas prioritariamente a seguinte ordem: bibliotecas digitais de universidades federais, bibliotecas digitais científicas e outras bibliotecas virtuais existentes na rede.

A Tabela 3.1 apresenta a lista de todas as bibliotecas digitais pesquisadas.

Tabela 3.1: *Bibliotecas digitais pesquisadas*

Ordem	Nome da Biblioteca Digital
1	<i>ACM Digital Library</i>
2	Biblioteca Digital de Teses e Dissertações da USP
3	Biblioteca Digital de Teses e Dissertações da Universidade Federal do Ceará
4	Biblioteca Digital de Teses e Dissertações - UFJF
5	Biblioteca Digital de Teses e Dissertações da Universidade Federal de Pelotas
6	Biblioteca Digital da Universidade Gama Filho
7	Biblioteca Digital da UFMG
8	Biblioteca Digital IBM <i>Online</i>
9	Biblioteca do Conhecimento
10	Biblioteca Nacional
11	Biblioteca Nacional Digital Brasil
12	Biblioteca Virtual de Direitos Humanos da Universidade de São Paulo - USP
13	Biblioteca Virtual de Educação
14	Biblioteca Virtual em Saúde
15	Bibliotecas Virtuais
16	CDIM - Centro de Documentação, Informação e Memória/UFG
17	CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
18	C@thedra - Biblioteca Digital de Teses e Dissertações
19	Comitê Gestor Internet/Brasil - Bibliotecas Virtuais
20	<i>Cornell University Library</i>
21	<i>Dialnet</i>
22	Divisão de Bibliotecas e Documentação da PUC-Rio
23	<i>EBSCO Publishing</i>
24	Escola do Futuro - Universidade de São Paulo/USP
25	<i>HighWire - Stanford University</i>
26	<i>IEEE Xplore Digital Library</i>
27	<i>ISI Web of Knowledge</i>
28	<i>Library of Congress</i>
29	LUME - Repositório Digital da Universidade Federal do Rio Grande do Sul
30	<i>MathSciNet - American Mathematical Society</i>
31	<i>Pilot Service (Engineering Repository Cross Search Demonstrator)/PerX</i>

32	Portal Domínio Público - Biblioteca digital desenvolvida em <i>software</i> livre
33	Repositório da Universidade de Lisboa
34	RepositóriUM
35	SciELO - Scientific Eletronic Library Online
36	<i>Scirus - for scientific information only</i>
37	Sistema de Bibliotecas da UFMG
38	Sistema de Bibliotecas da UNICAMP
39	Sistema Integrado das Bibliotecas da Universidade de Lisboa - SIBUL
40	<i>SpringerLink</i>
41	<i>Wiley - online library</i>

Na etapa de levantamento das bibliotecas digitais que usam o padrão *Dublin Core* para descrição de metadados, foi criado um catálogo identificando algumas informações extraídas desses repositórios: o nome da biblioteca digital, a url da página, a url de um registro como exemplo e os elementos do *Dublin Core* utilizados pelos repositórios.

Observa-se que a referência à forma simplificada e completa está relacionada à forma de visualização das informações disponibilizadas pelas bibliotecas digitais e não há uma relação direta quanto aos níveis Simples e Qualificado do *Dublin Core*. No caso das visualizações simplificadas, as bibliotecas digitais apresentam como informações mais comuns: título e autor do documento, palavras-chave, data de publicação, editora, resumo, *Uniform Resource Identifier* (uri); enquanto na visualização completa, além das informações comumente apresentadas pelas bibliotecas digitais, aparecem a linguagem definida para o documento, os direitos, o tipo do documento, o local da publicação e as pessoas que contribuíram para a publicação do conteúdo.

Dois exemplos dessa catalogação são apresentados na Tabela 3.2.

Tabela 3.2: Exemplos de metadados de duas bibliotecas digitais pesquisadas

Descrição	Conteúdo
Nome da Biblioteca Digital	Biblioteca Digital de Teses e Dissertações da USP
URL da página	http://www.teses.usp.br/
URL do registro (Exemplo)	http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29012001-160200/pt-br.php
Elementos (Simplificado)	documento, autor, nome completo, <i>e-mail</i> , unidade da USP, área do conhecimento, data de defesa, orientador, banca examinadora, título em português, palavras-chave em português, resumo em português, título em inglês, palavras-chave em inglês, resumo em inglês, arquivos, data de publicação, número de visitas, número de <i>downloads</i>
Elementos (Completo)	Não há
Descrição	Conteúdo
Nome da Biblioteca Digital	Repositório da Universidade de Lisboa
URL da página	http://repositorio.ul.pt/

URL do registro (Exemplo)	http://repositorio.ul.pt/handle/10451/1849
Elementos (Simplificado)	título, autor, palavras-chave, data, editora, resumo, uri, aparece nas coleções, ficheiro do registro (nome do documento, descrição, tamanho, formato (ver/abrir)).
Elementos (Completo)	degois.publication.location, degois.publication.title, dc.contributor.author, dc.contributor.author, dc.contributor.author, dc.date.accessioned, dc.date.available, dc.date.issued, dc.identifier.uri, dc.description.abstract, dc.language.iso, dc.publisher, dc.rights, dc.subject, dc.subject, dc.subject, dc.subject, dc.subject, dc.subject, dc.title, dc.type, aparece nas coleções

A partir do levantamento realizado foi executado um estudo sobre o quantitativo de ocorrências dos elementos utilizados por cada um dos repositórios. Certas bibliotecas digitais como o Repositório da Universidade de Lisboa utilizam tanto a forma simplificada de visualizar os dados quanto a forma completa para descrever os registros do repositório. Outras, como a Biblioteca do Conhecimento *Online*, Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, apresentam somente a forma simplificada para descrição das informações.

Nesta fase, os elementos similares foram agrupados para que a ocorrência fosse computada na mesma categoria. Os agrupamentos produzidos foram com os elementos título e *title*; autor, *author* e *creator*; resumo e *abstract*; uri e *link*; data e *date*; assunto e *subject*; editora, publicador e fornecedor; *e-mail*, endereço eletrônico; *location*, localização e local.

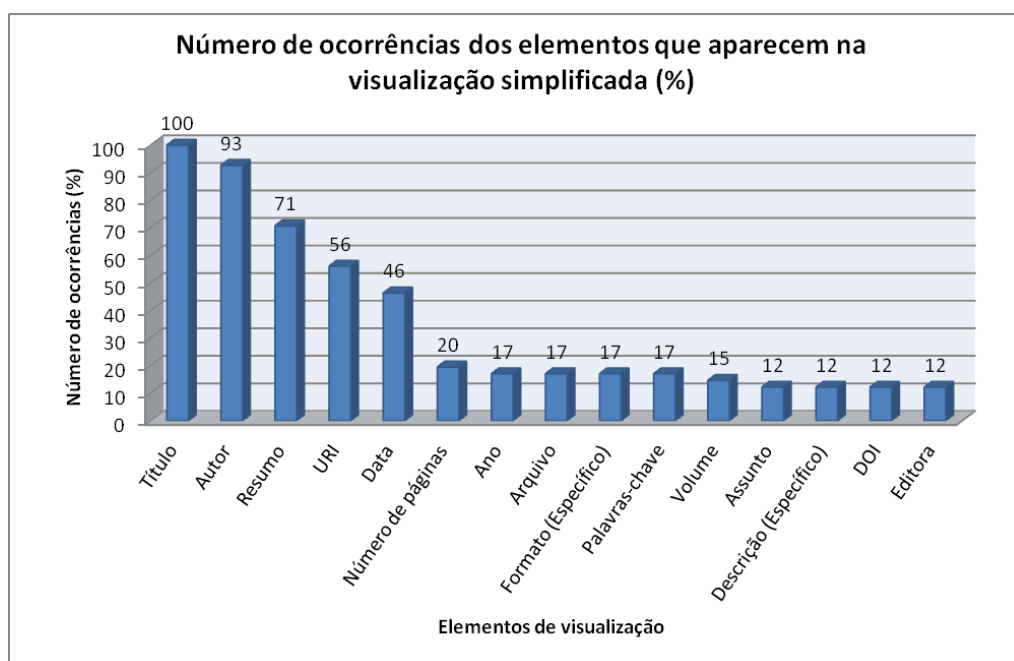


Figura 3.1: *Número de ocorrências dos elementos que aparecem na visualização simplificada*

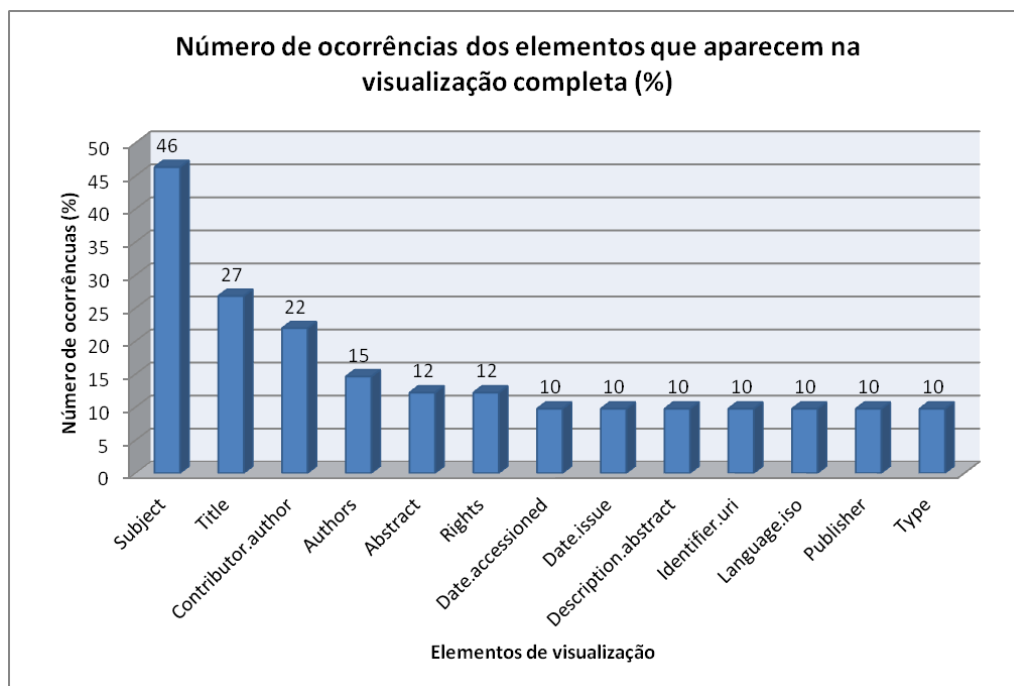


Figura 3.2: Número de ocorrências dos elementos que aparecem na visualização completa

A partir do levantamento pode-se observar que os elementos que tiveram mais ocorrências foram: título, autor, resumo, URI, data, número de páginas, ano, arquivo, formato, palavras-chave, volume, assunto, descrição, DOI e editora. Esses elementos referem-se a elementos na visualização simplificada, conforme é apresentado na Figura 3.1.

Os elementos na visualização completa que tiveram mais ocorrências foram: *subject*, *title*, *contributor.author*, *authors*, *abstract*, *rights*, *date.accessioned*, *date.issue*, *description*, *abstract*, *identifier.uri*, *language.iso*, *publisher* e *type*. Dentre as bibliotecas digitais que utilizam os elementos no formato completo, as representações desses elementos foram feitas em língua inglesa, seguindo a descrição do padrão *Dublin Core*; outras, como a Biblioteca Nacional Digital Brasil, utilizaram a língua portuguesa. O quantitativo dessas ocorrências pode ser visto na Figura 3.2.

Por se tratar de bibliotecas digitais observa-se que alguns elementos como DOI, volume e editora são campos próprios utilizados nesses repositórios.

A próxima subseção descreve como os elementos identificados são utilizados pelas bibliotecas digitais.

3.1.2 Significado dos Elementos do *Dublin Core* adotados pelas Bibliotecas Digitais

De acordo com a pesquisa realizada em bibliotecas digitais, pôde-se observar que os recursos do *Dublin Core* utilizados por elas seguem os conceitos introduzidos pelo padrão, sendo alguns deles ampliados ou especificados conforme a aplicação. Esses elementos são utilizados com os seguintes significados:

1. Elemento do *Dublin Core*: Título

Nomes similares: título, *title*, título em português, título em inglês;

Identificador: *title*;

Utilização: elemento utilizado para identificar o título do trabalho.

2. Elemento do *Dublin Core*: Criador

Nomes similares: autor, *author*, *creator*;

Identificador: *creator*;

Utilização: elemento utilizado para identificar o autor do trabalho.

3. Elemento do *Dublin Core*: Descrição

Nomes similares: resumo, *abstract*;

Identificador: *description*;

Utilização: elemento utilizado para descrever o resumo do trabalho em português ou inglês.

4. Elemento do *Dublin Core*: Identificador

Nomes similares: URI, *link*;

Identificador: *identifier*;

Utilização: elemento utilizado para citar ou criar um *link* para o trabalho.

5. Elemento do *Dublin Core*: Data

Nomes similares: data, *date*;

Identificador: *date*;

Utilização: elemento utilizado para informar a data de edição do trabalho.

6. Elemento do *Dublin Core*: Data

Nome: *data de publicação*;

Identificador: *date*;

Utilização: elemento utilizado para informar a data de publicação do trabalho.

7. Elemento do *Dublin Core*: Assunto

Nome: palavras-chave;

Identificador: *Subject*;

Utilização: elemento utilizado para descrever o conteúdo do recurso com palavras-chave.

8. Elemento do *Dublin Core*: Assunto

Nomes similares: assunto, *subject*;

Identificador: *Subject*;

Utilização: elemento utilizado para descrever o assunto do recurso.

9. Elemento do *Dublin Core*: Editor

Nomes similares: editora, publicador, fornecedor;

Identificador: *Publisher*;

Utilização: elemento utilizado para descrever o editor do recurso.

10. Elemento do *Dublin Core*: Criador

Nomes similares: *e-mail*, endereço eletrônico;

Identificador: *Creator*;

Utilização: elemento utilizado para descrever o endereço eletrônico do autor.

11. Elemento do *Dublin Core*: Fonte

Nomes similares: *location*, localização, local;

Identificador: *Source*;

Utilização: elemento utilizado para descrever o local onde o trabalho foi publicado (país, cidade). Em algumas bibliotecas digitais como a Dialnet a localização foi utilizada como fonte agrupando elementos como ISSN, volume, número, ano e páginas.

12. Elemento do *Dublin Core*: Fonte

Nome: fonte;

Identificador: *Source*;

Utilização: elemento utilizado para descrever um conjunto de informações como local onde o trabalho foi publicado (país, cidade), número da obra, ano de publicação e número de página.

3.2 Utilizando Elementos *Dublin Core* para descrever Metadados de Bancos de Dados Relacionais

Os elementos utilizados pelas bibliotecas digitais relacionados na Subseção 3.1.2 apesar de nomenclaturas diferentes recebem o mesmo conteúdo.

A nomenclatura utilizada para descrever metadados de bancos de dados relacionais terá como base a Tabela 2.3. Os nomes dos elementos foram definidos da seguinte forma: dc para indicar *Dublin Core* e sinal de *underline*(_), seguido do nome do elemento. Exemplo: dc_title. Já os elementos que não pertencem ao padrão *Dublin Core* são descritos de forma simples em língua inglesa: *provider*, *url*, *email*, *dispquery*, *loginbd* e *passwordbd*.

Como resultado final de tal classificação de ocorrência, chegou-se a um conjunto de 23 atributos que devem ser considerados para exportar os metadados referentes ao contexto de bibliotecas digitais. A Tabela 3.3 resume o conjunto de atributos utilizados para descrever os metadados de um banco de dados. São eles:

Tabela 3.3: Atributos da tabela de metadados para exposição (TME)

serial	provider	url	email	oai_set	dispquery
dc_title	dc_creator	dc_subject	dc_description	dc_contributor	dc_publisher
dc_date	dc_type	dc_format	dc_identifier	dc_source	dc_language
dc_relation	dc_coverage	dc_rights	loginbd	passwordbd	-

Os elementos do *Dublin Core* selecionados receberão os seguintes conteúdos:

- **dc_title:** utilizado para apresentar o título do banco de dados. O campo título é utilizado conforme descrito no *Dublin Core*, o nome como o recurso é formalmente conhecido [26];
- **dc_creator:** utilizado para informar o nome do autor ou entidade responsável pelo recurso;
- **dc_subject:** na recomendação do *Dublin Core*, o *subject* deverá ser representado usando palavras-chave, frases-chave ou código de classificação. O *Dublin Core* recomenda ainda que a melhor prática é usar vocabulários controlados [26]. É utilizado para informar as palavras-chave associadas ao conteúdo do banco de dados;

- **dc_description:** utilizado para apresentar um resumo (textual) sobre o conteúdo do banco de dados;
- **dc_contributor:** utilizado para informar o nome da pessoa/instituição responsável por qualquer contribuição para o conteúdo do recurso;
- **dc_publisher:** utilizado para informar o nome da entidade/instituição responsável por tornar o recurso acessível;
- **dc_date:** utilizado para informar a data da publicação do recurso;
- **dc_type:** utilizado para informar o gerenciador adotado para implementar o banco de dados com o valor padrão de "database" seguido pelo seu tipo. Ex.: *MySQL*, *PostgreSQL*;
- **dc_format:** utilizado para descrever o formato do banco de dados com o valor padrão de "SGBDR";
- **dc_identifier:** utilizado para descrever o nome do banco de dados conforme criado pelo administrador desse banco;
- **dc_source:** utilizado para informar o principal departamento que faz uso do banco de dados;
- **dc_language:** utilizado para informar o idioma que descreve o recurso;
- **dc_relation:** utilizado para relacionar os sistemas que utilizam o banco de dados;
- **dc_coverage:** utilizado para descrever as áreas de interesse do banco de dados (financeira, cadastro, administrativo, vendas, acadêmico etc.);
- **dc_rights:** utilizado para registrar os direitos sobre o recurso.

Além dos elementos *Dublin Core* já relacionados anteriormente, foram criados alguns elementos complementares para o armazenamento de dados não contemplados pelos elementos básicos do padrão *Dublin Core*. Esses elementos estão destacados com o símbolo asterisco (*) na Tabela 3.4. Os elementos destacados com duplo asterisco

(**), além de serem campos complementares, indicam a necessidade da autorização do administrador do banco de dados para acessar o conteúdo do repositório.

- **serial**: utilizado para gerar uma sequência do repositório cadastrado;
- **oai_set**: utilizado para informar a url do sítio ou endereço onde o sistema está hospedado. Este campo é utilizado no relatório do sistema para direcionar para a url da página;
- **provider**: utilizado para informar o endereço do provedor de dados do banco de dados;
- **url**: utilizado para informar a url ou endereço do sítio vinculado ao banco de dados;
- **email**: utilizado para informar o endereço eletrônico do administrador do banco de dados;
- **dispquery**: este campo é um *flag* que indica que o conteúdo do banco de dados está disponível para consulta na *web*.

Os dois elementos a seguir foram criados para acesso a banco de dados.

- **loginbd**: usuário do banco de dados utilizado para acessar o banco de dados;
- **passwordbd**: senha do banco de dados utilizada para acessar o banco de dados.

Os elementos *loginbd* e *passwordbd* são aqui destacados separadamente para indicar que são informações as quais necessitam de privilégios especiais e da autorização do administrador do banco de dados.

A estrutura desses elementos e alguns campos adicionais criados para descrever os metadados de banco de dados são definidos na Seção 4.5.2 da Tabela 4.2.

A mesma estrutura foi utilizada para armazenar as informações de bibliotecas digitais, porém os elementos *url*, *dc_identifier*, *dc_source*, *dc_language*, *dc_relation*, *loginbd* e *passwordbd* definidos para os bancos de dados não serão utilizados para descrever as bibliotecas digitais. Para diferenciar os conteúdos de bancos de dados de bibliotecas digitais são utilizados os elementos *dc_type* e *dc_format*.

A Subseção 3.2.1 apresenta um exemplo de uso da estrutura proposta.

3.2.1 Exemplo de Uso da Estrutura Proposta

Para exemplificar o preenchimento do conteúdo da estrutura proposta na Seção 4.2, foi utilizado um banco de dados fictício de produtores rurais. A descrição do banco de dados, que descreve os metadados a serem expostos, pode ser vista com detalhes na Tabela 3.4.

Tabela 3.4: Exemplo de uso da estrutura dos elementos do *Dublin Core* e elementos adicionais para descrição de metadados de um banco de dados fictício

Elemento	Exemplo
serial*	1
provider*	http://localhost/website/xyzsystem/database/oai
url*	http://localhost/website/xyzsystem/
email*	admname@xyz.com
oai_set	http://localhost/website/system/systemname
dispquery**	não
dc_title	Database xyz
dc_creator	XYZ
dc_subject	Plantio de flores e verduras. Terra. Sustentabilidade.
dc_description	O plantio de flores e verduras são fontes de economia para grande parte das famílias rurais, que trabalham a terra e a utilizam de forma sustentável.
dc_contributor	XYZ
dc_publisher	<i>Rose Publisher</i>
dc_date	2011-01-01
dc_type	Database - MySQL
dc_format	SGBDR
dc_identifier	bdprodutoresrurais
dc_source	Sindicato Rural WXP
dc_language	Português
dc_relation	Sistema de Cadastro de Produtores Rurais
dc_coverage	Produtores Rurais do Município XSM
dc_rights	Sindicato Rural WXP
loginbd**	*****
passwordbd**	*****

(*) Elementos adicionais; (**) Elementos adicionais que requerem autorização do administrador do banco de dados.

Como pode ser observado na Tabela 3.4, os elementos propostos nesta seção para descrever as informações dos bancos dados foram preenchidos conforme a interpretação do elemento no contexto de banco de dados.

O elemento *provider* deve conter obrigatoriamente o radical <oai_set/database/oai>, seguido do nome do arquivo XML, em que:

- *oai_set*: indica a página *web* onde se encontra o sítio do aplicativo;
- *database*: indica a fonte de dados de um banco de dados;
- *oai*: indica que é um arquivo aberto para consulta;
- *<nome-do-arquivo>*: nome do arquivo XML gerado para disponibilizar a descrição do banco de dados.

No exemplo apresentado observa-se que o campo *dispquery* está preenchido com o valor "não". Isso indica que o conteúdo do banco de dados não está disponível na rede para consulta. A utilização deste *flag* indica que o usuário poderá ou não realizar consulta a banco de dados. Os elementos *dc_subject* e *dc_date* são de grande relevância, já que as informações poderão ser recuperadas a partir das palavras-chave e da data de publicação informadas.

Os elementos *dc_identifier*, *loginbd* e *passwordbd* permitirão o acesso direto ao banco de dados, possibilitando que os dados possam ser recuperados a partir da solicitação de requisição do usuário.

A Subseção 3.2.2 apresenta uma descrição de um mecanismo para interoperabilidade de banco de dados com outras fontes de informação.

3.2.2 Um Mecanismo para Interoperabilidade de Banco de Dados com outras Fontes de Informação

Nesta Seção é apresentada uma proposta de um mecanismo que permite o compartilhamento de informações armazenadas em bancos de dados relacionais com outras fontes de informação, neste texto, exemplificadas por bibliotecas digitais.

O mecanismo proposto é executado em três etapas:

1. Descrição de metadados;
2. Extração de metadados;
3. Coleta e exposição de metadados.

A Figura 3.3 ilustra o contexto do mecanismo proposto:

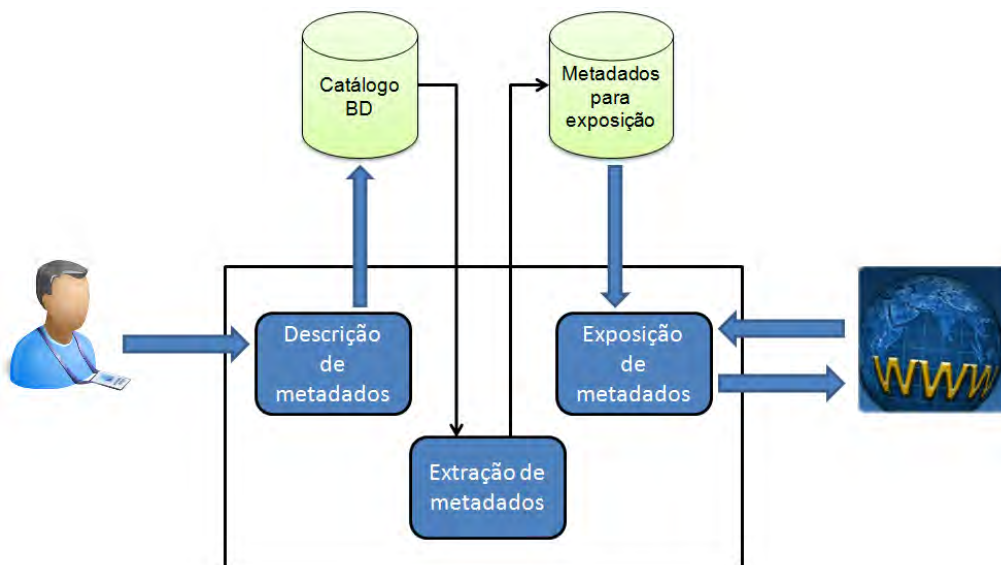


Figura 3.3: Modelo do extrator de metadados proposto

Cada uma dessas etapas é descrita a seguir, e os passos para a execução do mecanismo proposto encontram-se representados na Figura 3.4.

1. Descrição de metadados

Conforme discutido no capítulo 2, os metadados utilizados para descrever um recurso de informação em repositórios digitais não estão disponíveis nos catálogos de bancos de dados. Por isso, a primeira etapa do mecanismo proposto sugere que os metadados de bancos de dados sejam fornecidos pelo administrador de bancos de dados, detentor dos direitos de acesso e conhecedor do conteúdo desse repositório, e armazenados em uma tabela que será uma extensão do catálogo do sistema gerenciador de banco de dados.



Figura 3.4: Passos para a execução do mecanismo proposto

Essa tabela para armazenamento de metadados, denominada tabela de metadados

para exposição (TME), contém um atributo para cada um dos elementos *Dublin Core* escolhidos para representar os metadados para exportação de um banco de dados.

Com a realização da descrição dos metadados, a etapa seguinte trata da extração de metadados informados na extensão do catálogo.

2. Extração de metadados

Nesta etapa, é realizado o processo de extração dos metadados armazenados na tabela descrita no item anterior. Consiste em: (1) consultar a tabela de metadados; (2) extrair os metadados relacionados a uma determinada fonte de informação e (3) colocar o conjunto de metadados extraídos em um formato apropriado para exposição.

A extração de metadados poderá ser realizada de duas formas: (1) gerar metadados de um banco de dados ou (2) gerar metadados de todos os bancos de dados que foram cadastrados. Em ambas as formas é gerado um arquivo XML, que será armazenado em um repositório; posteriormente ele poderá ser acessado para recuperação dos metadados expostos.

A escolha na forma de gerar os metadados a serem expostos segue um critério diretamente relacionado ao vínculo da aplicação com o banco de dados. Se a aplicação possuir uma URL e um banco de dados vinculado a essa URL, então os metadados serão gerados individualmente para aquele banco de dados. A URL do aplicativo manterá um repositório indicado por `/database/oai`, que indica a localização do repositório de metadados e descreve o banco de dados daquela aplicação.

Os bancos de dados que não atenderem ao critério anteriormente descrito serão agrupados, e os metadados serão gerados em um único arquivo XML. Esse arquivo XML deverá ser disponibilizado na URL do aplicativo principal da organização.

3. Coleta e exposição de metadados

A terceira etapa do mecanismo proposto neste trabalho é responsável pela coleta dos metadados que estão armazenados no repositório de metadados para exposição (RME), disponibilizando-os para serem recuperados pelos usuários da *internet*.

Uma vez que os metadados estejam armazenados em um repositório, estes estão disponíveis para serem recuperados pelos usuários da *internet*. Neste sentido, esta terceira etapa do mecanismo proposto atua como um provedor de dados que expõem metadados por meio do protocolo OAI-PMH. A execução dessa etapa ocorre quando um provedor de serviço (*harvesters*) faz uma solicitação de serviços OAI-PMH para a coleta de metadados. Em seguida, o repositório de metadados é acessado para dar uma resposta

adequada à consulta de acordo com os verbos de requisição¹ do protocolo (ver Seção 2.4). A coleta e a exposição de metadados serão realizados por intermédio de uma interface de consulta que será disponibilizada aos usuários.

Assim, foram descritas as três etapas do mecanismo proposto para interoperabilidade de banco de dados com outras fontes de informação. Para demonstrar a viabilidade prática do mecanismo aqui proposto, a próxima seção descreve um sistema (protótipo) desenvolvido para implementar a abordagem descrita.

¹Os verbos de requisição são comandos transmitidos aos repositórios do protocolo OAI-PMH e que por sua vez retornam as respostas codificadas em XML. Os verbos de requisição são: *Identify*, *ListMetadataFormats*, *ListSets*, *ListIdentifiers*, *ListRecords* e *GetRecord*.

Um Sistema de Extração e Exposição dos Metadados de Banco de Dados Relacionais

Existe por trás da mente de todo artista um padrão ou tipo de arquitetura. G. K. Chesterton.

Gastamos muito tempo - a maioria do esforço do projeto - não implementando ou testando, mas tentando decidir o que construir. Brian Lawrence.

O sistema de extração e exposição de metadados de bancos de dados relacionais, denominado *METADATABASE EXTRACTOR*, é uma solução proposta para disponibilizar os metadados de bancos de dados relacionais e tornar possível a interoperabilidade destes com outros repositórios disponíveis na *web*. Este capítulo define os objetivos e os problemas que delimitam o escopo do *METADATABASE EXTRACTOR*, bem como os requisitos de *software*.

Na sequência são apresentados os objetivos definidos para o *METADATABASE EXTRACTOR* (Seção 4.1), os principais problemas encontrados durante o trabalho e, a partir desses problemas, são definidas as necessidades e o escopo do sistema (Seção 4.2). Os principais atores são descritos na Seção 4.3, seguidos dos requisitos do sistema (Seção 4.4), a implementação do protótipo (Seção 4.5) e a apresentação da arquitetura do sistema (Seção 4.5.5).

4.1 Objetivos

Diante do contexto que se pretende atender (Seção 1.1) e baseado nos trabalhos pertinentes à área de estudo (Seção 2.1), os objetivos da proposta de interoperabilidade de bancos de dados relacionais com outras fontes, utilizando o sistema *METADATABASE EXTRACTOR* incluem:

- **contribuição na área (O1):** auxiliar pesquisas na área de interoperabilidade de bancos de dados com outros repositórios da *web* facilitando a experimentação de propostas para recuperação de informações de bancos de dados relacionais;
- **contemplar abordagem recorrente (O2):** contemplar uma abordagem comumente utilizada em bibliotecas digitais, o que assegura uma maior aplicabilidade para bancos de dados relacionais;
- **facilidade de uso (O3):** a proposta deve ser "fácil" de ser utilizada;
- **facilidade de gerência (O4):** a solução deve facilitar a disponibilização e a coleta de dados;
- **aplicações originais(O5):** a solução deve favorecer a interoperabilidade, sem contudo exigir alterações nas fontes de dados, neste caso, dos bancos de dados relacionais. Assim, não é preciso ter acesso aos bancos de dados para extrair metadados, já que os metadados serão fornecidos pelos administradores de bancos de dados.

4.2 Principais Problemas, Necessidades e Escopo

Os principais problemas levantados durante o desenvolvimento da pesquisa foram:

- Informações sobre um determinado assunto podem estar armazenadas em diferentes repositórios: banco de dados, bibliotecas digitais, páginas *web* etc.;
- Uma única consulta de usuário pode acarretar uma pesquisa em diferentes fontes de dados;
- Como favorecer a interoperabilidade de dados armazenados em bancos de dados com dados provenientes de outras fontes?

Ao considerar os problemas especificados acima, definiram-se as necessidades que o sistema deve atender:

- Disponibilizar informações de bancos de dados na *web*;
- Descrever as informações dos bancos de dados utilizando o padrão *Dublin Core*;
- Expor os metadados por meio do protocolo OAI-PMH;
- Realizar consulta a banco de dados a partir de palavras-chave.

O objetivo do projeto é desenvolver um protótipo de um extrator de metadados que descreva os metadados extraídos do banco de dados relacionais nos elementos do padrão *Dublin Core* e exponha esses metadados por meio do protocolo OAI-PMH para serem consultados na *internet*.

As funcionalidades previstas no sistema deverão ser contempladas, verificando-se o cronograma previsto para o seu desenvolvimento e as tecnologias disponíveis para a

sua implementação. As tecnologias utilizadas para implementar o aplicativo deverão ser, preferencialmente *open source* e serão utilizadas a linguagem de programação *Hypertext Preprocessor* (PHP), Javascript, *Hypertext Markup Language* (HTML), *Cascading Style Sheets* (CSS) e algumas bibliotecas disponíveis.

O aplicativo deverá contemplar as seguintes características: clareza, facilidade na utilização, cores neutras/ leves e ter *layout* padrão para as funcionalidades implementadas. A escolha das cores e das fontes de letras deverá ser feita por meio do padrão adotado e definido no documento.

O *layout* principal, de preferência, deve incorporar *links* que serão apresentados em forma de página *web*, trazendo textos explicativos como uma breve apresentação, objetivos, motivação e proposta do sistema, além da possibilidade de agregar futuros melhoramentos.

O projeto deverá contemplar as funcionalidades básicas definidas no escopo. Caso haja alteração durante o seu desenvolvimento, deverá ser documentado, apresentando-se as justificativas da mudança.

4.3 Principais Atores

Os atores na engenharia de *software* são as diferentes pessoas (ou dispositivos) que usam o sistema ou produto dentro do contexto da função e do comportamento que devem ser descritos [58]. A seguir são descritos os atores que interagirão com o sistema:

- **administrador de banco de dados:** profissional com experiência e conhecimento em administração de bancos de dados. Seu papel é manter as informações dos bancos de dados atualizadas no sistema;
- **usuário de sistema da organização:** profissional com conhecimento de sistemas e de bancos de dados legados da organização. Este usuário poderá ocupar o lugar do administrador de banco de dados caso esse profissional não exista em uma organização;
- **desenvolvedor:** profissional com conhecimento em desenvolvimento de sistemas e arquitetura *web*. Seu principal objetivo nesse contexto é manter o sistema funcionando na *web* e fazer adaptação no sistema caso necessário;
- **usuários de sistema:** qualquer profissional da organização que necessite realizar consultas no banco de dados.

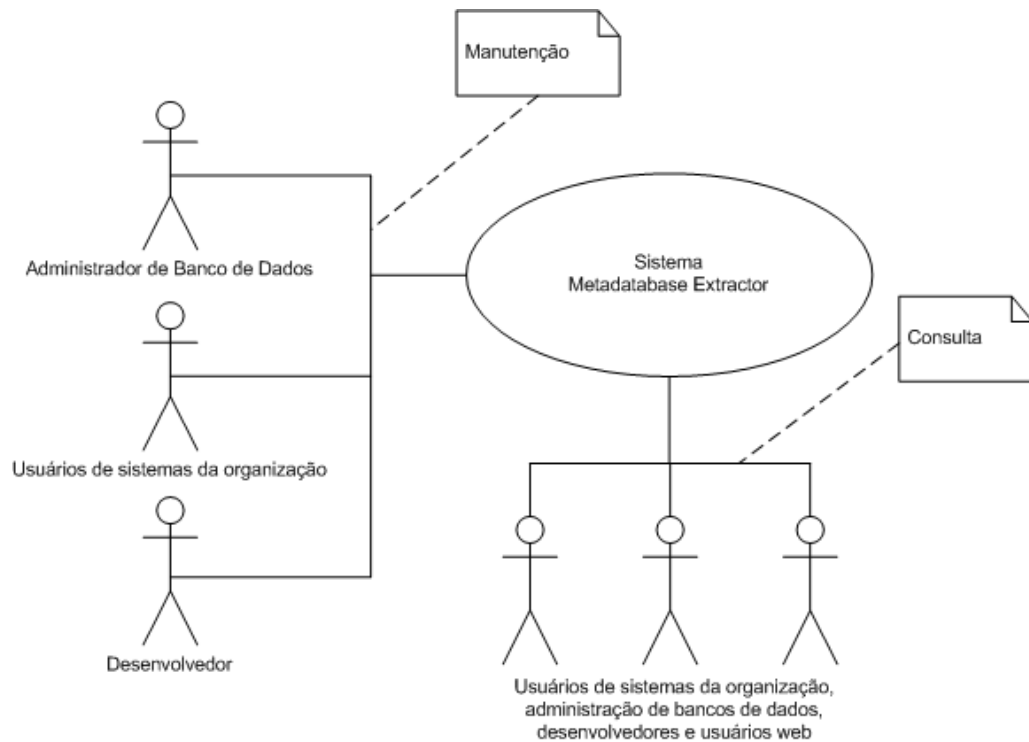


Figura 4.1: *Classificação dos atores quanto ao uso da solução do sistema*

4.4 Requisitos de *Software*

A engenharia de requisitos começa com a fase de concepção, tarefa que define o escopo e a natureza do problema a serem resolvidos. Para a especificação do sistema foi utilizada a pesquisa de levantamento do cenário na área por meio de artigos científicos e conversas informais com professores do grupo de pesquisa do Instituto de Informática da UFG, o que possibilitou o desenvolvimento da pesquisa.

Após o entendimento dos problemas e as necessidades que o sistema deve atender, foram relacionados os requisitos a ele direcionados. Os requisitos do sistema, são apresentados na Tabela 4.1.

Tabela 4.1: *Requisitos do software*

Identificação do requisito	Descrição dos requisitos
R1	O administrador de banco de dados/sistema informa os metadados de um determinado banco de dados relacional.
R2	O sistema tem acesso aos metadados de banco de dados independentemente do tipo de banco de dados relacional.

R3	O administrador de banco de dados poderá alterar ou complementar posteriormente os metadados extraídos.
R4	O administrador do banco de dados deverá informar o banco de dados que deseja disponibilizar.
R5	O sistema descreve os metadados extraídos e os configura no padrão <i>Dublin Core</i> .
R6	O sistema gera um arquivo XML dos metadados extraídos.
R7	O sistema analisa o arquivo XML gerado e valida o arquivo.
R8	Se o arquivo gerado não for válido, é solicitada uma nova requisição para gerar o arquivo com a solicitação de correção de erros.
R9	O sistema disponibiliza os arquivos gerados para o servidor de dados por meio do protocolo OAI-PMH.
R10	O sistema disponibiliza consulta aos metadados expostos.
R11	O usuário realiza diferentes consultas à base de dados exportada.
R12	O sistema retorna a pesquisa ao usuário de acordo com a solicitação realizada.

4.5 Implementação do Protótipo

Nesta seção estão descritas as tecnologias empregadas para a implementação do protótipo (Seção 4.5.1) e a estrutura da Tabela de Metadados para Exposição (TME) (Seção 4.5.2). Na (Seção 4.5.3) é apresentado um modelo de solução proposto para o sistema, além de estender as visões para os usuários de sistemas (Seção 4.5.4). A arquitetura do sistema é descrita na (Seção 4.5.5) seguido da descrição e da funcionalidade do sistema (Seção 4.5.6).

4.5.1 Tecnologias Empregadas

Para o desenvolvimento do *METADATABASE EXTRACTOR* foram utilizados os seguintes *softwares*: PHP 5.2.6, Apache 2.0.63, Mysql 5.0.51b, PhpMyAdmin 2.11.7, XML, Macromedia Dreamweaver MX, Macromedia Fireworks MX, Mozilla Firefox, Filezilla Client e Sistema Operacional Windows 7 Home Premium (também foram efetuados testes em um Servidor Linux).

4.5.2 Estrutura da Tabela de Metadados para Exposição (TME)

Os dados de bancos de dados e de bibliotecas digitais testados foram armazenados na seguinte estrutura:

Tabela 4.2: *Estrutura dos elementos do Dublin Core e elementos adicionais para descrição dos metadados de banco de dados*

Elemento	Tipo da variável	Tamanho	Interpretação do elemento no contexto de banco de dados
serial*	inteiro	11	Indicador de sequência do registro armazenado na tabela
provider*	string	255	URL do provedor de dados do banco de dados; local onde o arquivo XML estará disponível para consulta. Localização indicada por: <oai_set/database/oai/<nome-do-arquivo.xml>
url*	string	255	URL ou endereço do sítio vinculado ao banco de dados
email*	string	255	Endereço eletrônico do administrador do banco de dados
oai_set	string	255	URL do sítio ou endereço onde o sistema está hospedado
dispquery **	boolean	5	Flag que indica a disponibilização do conteúdo na rede
dc_title	string	255	Título do banco de dados
dc_creator	string	255	Nome do autor ou entidade responsável pelo recurso
dc_subject	string	255	Palavras-chave referente ao banco de dados
dc_description	text		Resumo do banco de dados
dc_contributor	string	255	Nome da pessoa/instituição responsável por qualquer contribuição para o conteúdo do recurso
dc_publisher	string	255	Nome da entidade/instituição responsável por tornar o recurso acessível
dc_date	date	10	Data da publicação do recurso. Formato YYYY-MM-DD
dc_type	string	255	Tipo do banco de dados com o valor padrão de "database" seguido pelo seu tipo. Ex.: MySQL, PostGreSQL
dc_format	string	255	Formato de banco de dados com o valor padrão de "SGBDR"
dc_identifier	string	255	Nome do banco de dados conforme criado pelo administrador do banco de dados

dc_source	string	255	Principal departamento que utiliza o banco de dados
dc_language	string	255	Língua utilizada para descrever o recurso
dc_relation	string	255	Sistemas que utilizam o banco de dados
dc_coverage	string	255	Cobertura do banco de dados (financeira, cadastro, administrativo, vendas, acadêmico etc.)
dc_rights	string	255	Informação de direitos sobre o recurso
loginbd**	string	15	<i>Login</i> para acesso ao banco de dados
passwordbd**	string	15	Senha de acesso ao banco de dados

(*) Elementos adicionais; (**) Elementos adicionais que requerem autorização do administrador do banco de dados.

Os elementos destacados com o símbolo asterisco (*) na Tabela 4.2 são elementos complementares para o armazenamento de dados não contemplados pelos elementos básicos do padrão *Dublin Core*. Os elementos destacados com duplo asterisco (**), além de serem campos complementares, indicam campos para acessar o banco de dados e necessitam de autorização do administrador do banco de dados por envolver segurança de dados. Os administradores de dados podem ou não informar os conteúdos desses campos.

4.5.3 Modelo de Solução Proposto

A partir do fluxo de execução de um repositório digital apresentado na Figura 2.6, foi desenhado um fluxo de extração e exposição de metadados de bancos de dados relacionais. Esse fluxo pode ser visto na Figura 4.2.

O fluxo apresentado na Seção 2.5 (Figura 2.6) foi adaptado para o contexto deste trabalho. As fontes de informações antes indicadas por livros, mapas, sons, material gráfico etc. foram adaptadas para os bancos de dados. Diferentemente da forma que o repositório digital realiza o processo de seleção dos materiais, por meio de aquisições, o processo de extração de metadados de bancos de dados é realizado a partir do fornecimento das informações pelo administrador desses bancos, o qual detém conhecimento sobre o conteúdo do repositório.

Enquanto os serviços técnicos em repositórios digitais envolvem verificação do conteúdo, acréscimo de metadados, licenciamento do registro, direitos de administração e controle de acesso; em bancos de dados os serviços estão voltados para descrição do conteúdo por meio do padrão *Dublin Core* geração de arquivo de metadados, análise do arquivo gerado, exposição dos metadados e direitos de administração e controle de acesso comuns aos repositórios digitais.

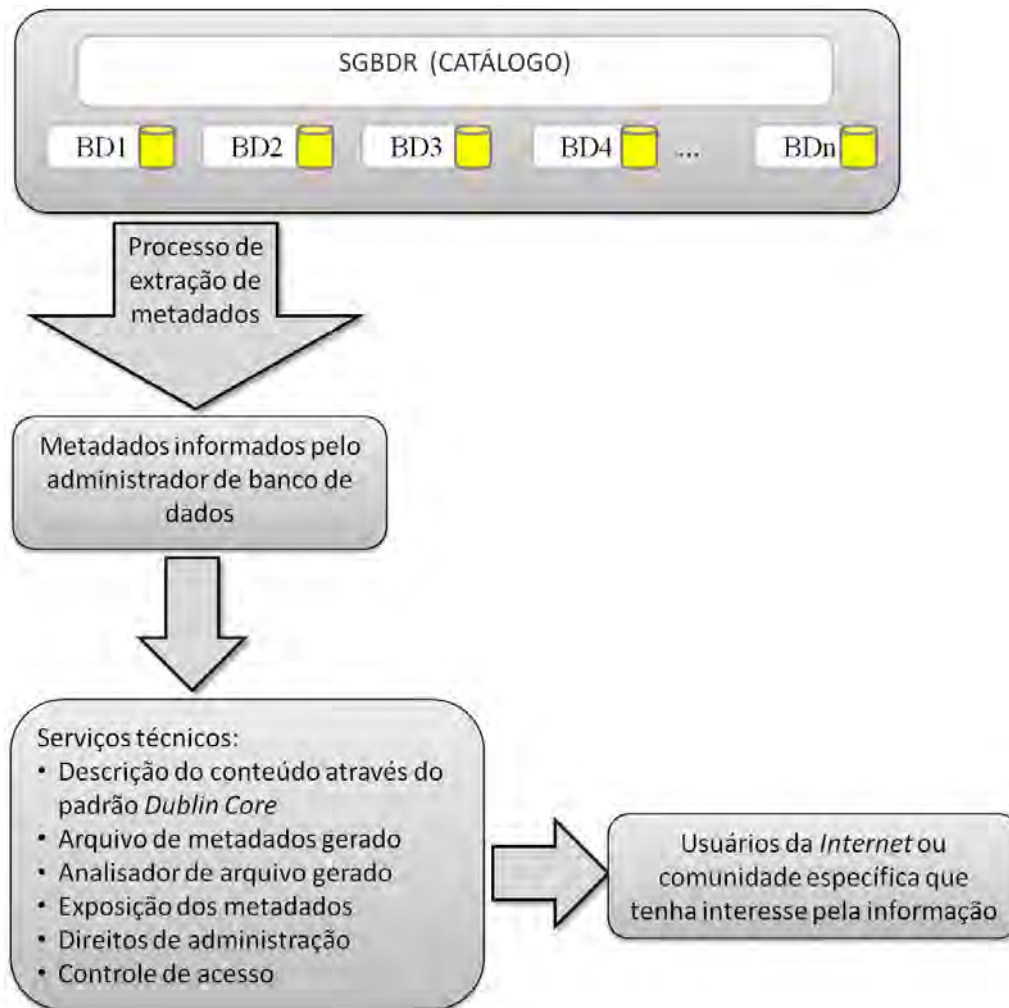


Figura 4.2: Fluxo de extração e exposição de metadados de bancos de dados relacionais. Adaptado de [35]

O fluxo final em repositórios digitais termina quando os serviços técnicos são concluídos, e o material disponibilizado aos usuários da biblioteca digital. No caso de metadados de banco de dados, as descrições são disponibilizadas aos usuários da *internet* ou a uma comunidade específica interessada pela informação.

A partir desse fluxo foi criado um modelo de solução apresentado na Figura 4.3.

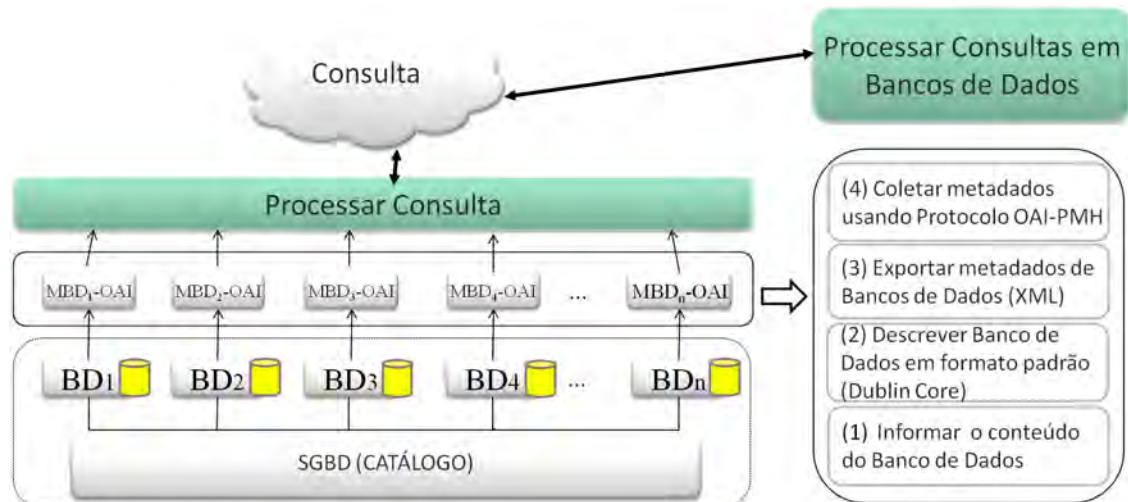


Figura 4.3: Modelo de solução do sistema proposto

O modelo de solução destaca quatro passos até a consulta contextualizada de informações em bancos de dados:

1. **Informar o conteúdo do Banco de Dados:** o administrador do banco de dados deverá informar sobre o conteúdo do banco de dados, já que os catálogos desses bancos não descrevem detalhes do repositório;
2. **Descrever o Banco de Dados em formato padrão *Dublin Core*:** será utilizado para descrever os metadados de bancos de dados;
3. **Exportar metadados de Bancos de Dados (XML):** os metadados extraídos e descritos no padrão *Dublin Core* serão apresentados no arquivo XML (uma linguagem que permite a construção de documentos legíveis para seres humanos e que podem ser facilmente tratados por máquinas [66]);
4. **Coletar e expor metadados usando Protocolo OAI-PMH:** o protocolo será utilizado para compartilhar metadados entre os participantes da iniciativa.

4.5.4 Visões dos Usuários do Sistema

As funcionalidades do sistema são disponibilizadas conforme o tipo de acesso do usuário. Portanto, o sistema apresenta duas visões: visão do usuário comum e visão do administrador de sistemas.

1. **Visão do usuário comum:** permite acesso aos *links*: principal, objetivo, motivação, proposta do projeto, coleta de metadados, sistema de busca, referência e informações sobre o sistema (com exceção da opção administração do sistema) conforme pode ser visto na Figura 4.6.
2. **Visão do administrador de sistema:** o administrador do sistema tem acesso a todas as funcionalidades do sistema, principalmente à interface de gerenciamento do sistema.

4.5.5 Arquitetura do Sistema

Para Graves (2003, p. 180) [29], a arquitetura é a organização da funcionalidade do sistema em módulos interconectados, e cada módulo é um componente do sistema que desempenha várias funções relacionadas. O objetivo de uma boa arquitetura de sistema é agrupar apropriadamente requisitos funcionais em módulos e organizar esses módulos em um sistema que atenda a todos os requisitos.

Já para Bass, Clements e Kazman (*apud* PRESSMAN 2006, p. 208) [58], a arquitetura de *software* de um programa ou sistema computacional é a estrutura ou estruturas do sistema que abrangem os componentes de *software*, as propriedades externamente visíveis desses componentes e as relações entre eles.

A partir dessas definições, a proposta desta arquitetura abrangerá questões como:

- Quantos módulos o sistema conterà?
- Como esses módulos serão conectados?
- Qual será a funcionalidade de cada módulo?

A ideia do *METADATABASE EXTRACTOR* (extrator de metadados) é disponibilizar os metadados extraídos de diferentes bancos de dados, possibilitando que suas informações possam ser conhecidas externamente como acontece com as informações disponibilizadas pelas bibliotecas digitais, por exemplo.

O sistema tem como entrada os recursos de informação, aqui representados por banco de dados relacional, não modificando os dados da fonte, mas apenas utilizando-os para descrever os metadados extraídos dessas fontes e disponibilizá-los em um provedor de dados para consulta.

A Figura 4.4 apresenta uma visão da arquitetura do sistema "*METADATABASE EXTRACTOR*" proposto:

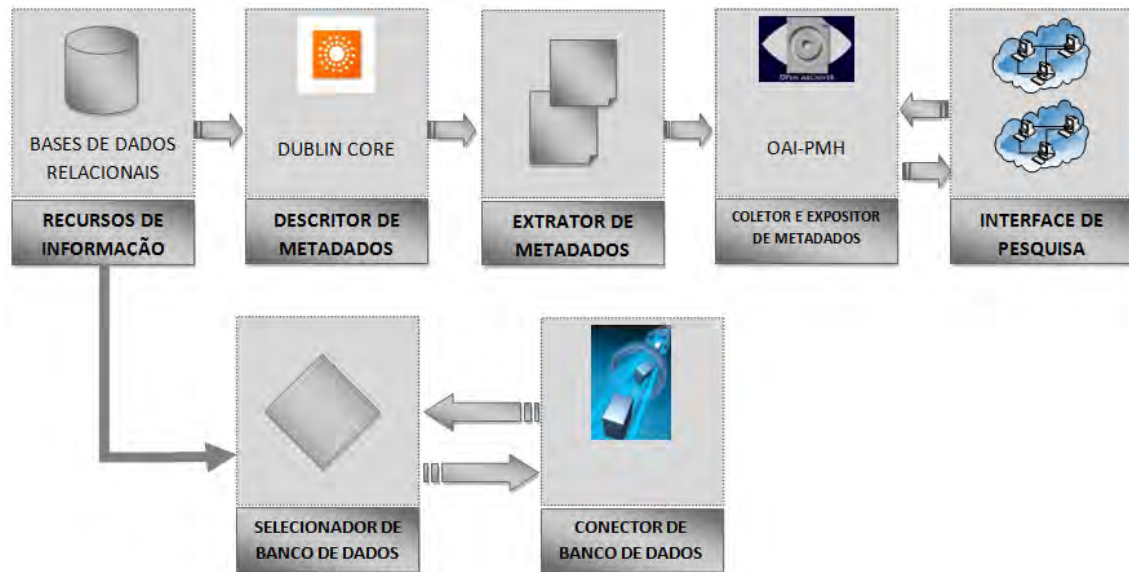


Figura 4.4: Arquitetura do sistema proposto de extração de metadados (METADATABASE EXTRACTOR)

Esta arquitetura pode ser descrita da seguinte forma:

1. **Camada de recursos de informação:** composto pelo recurso de informação representado aqui por um banco de dados relacional;
2. **Selecionador de banco de dados:** apresenta uma lista de banco de dados pré-configuradas e é responsável pela configuração dos parâmetros para conexão com o banco de dados no servidor;
3. **Conector com o banco de dados:** responsável pela conexão com o banco de dados selecionado;
4. **Extrator de metadados:** extrai os metadados do banco de dados selecionado;
5. **Descritor de metadados:** descreve os metadados extraídos e os configura no padrão *Dublin Core*;
6. **Coletor e expositor de metadados:** disponibiliza os arquivos gerados para o servidor de dados por meio do protocolo OAI-PMH;
7. **Interface de consulta:** responsável pela apresentação dos resultados de uma consulta; está disponível para outros provedores de serviços.

A arquitetura de conteúdo adotada para o protótipo foi a estrutura hierárquica como pode ser visto na Figura 4.5. Estruturas hierárquicas são a arquitetura mais comum de Aplicações baseadas na *Web (WebApp)* e pode ser projetada de maneira que permita um fluxo de controle horizontalmente (por meio da ramificação de hipertexto) ao longo dos ramos verticais da estrutura. Assim, o conteúdo apresentado no ramo mais à esquerda da hierarquia pode ter ligações de hipertexto que levem ao conteúdo existente no meio, ou no ramo mais à direita da estrutura [58].

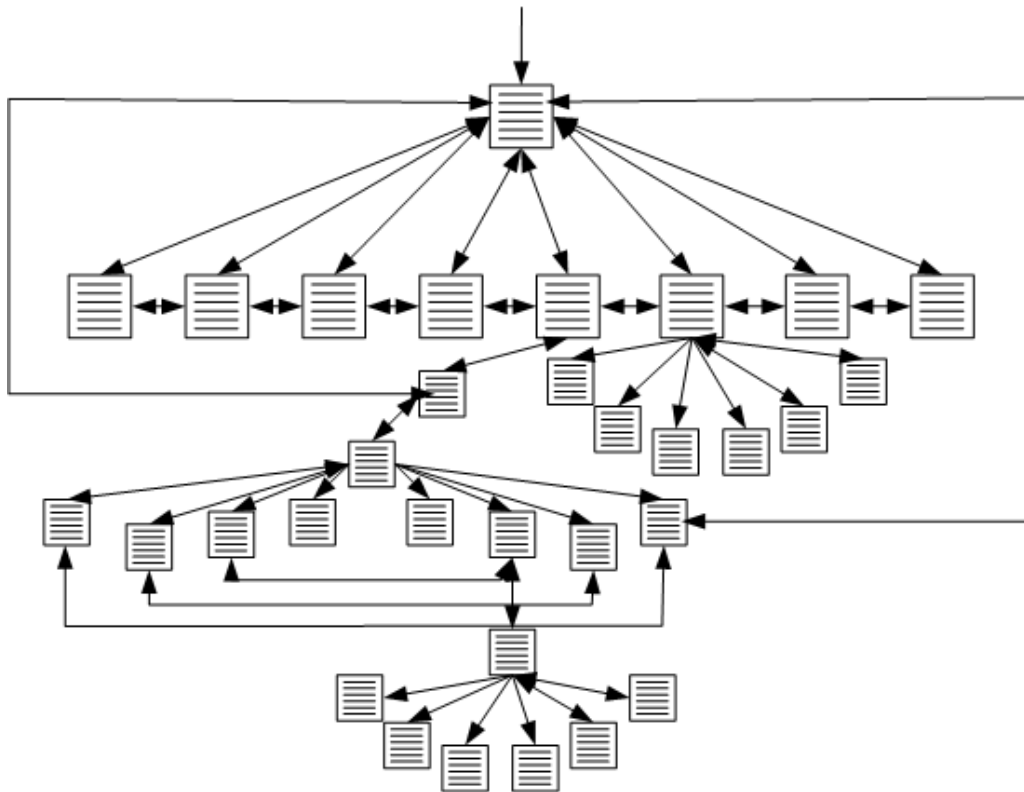


Figura 4.5: *Arquitetura de conteúdo adotada para o protótipo desenvolvido. Adaptado de [58]*

4.5.6 Descrição do Protótipo

O protótipo foi desenvolvido tendo em vista a utilização em ambientes *web*. Ao acessar o aplicativo no navegador, será apresentada ao usuário uma página com os *links*: principal, objetivo, motivação, proposta do projeto, administração do sistema, coleta de metadados, sistema de busca, referências e informações sobre o sistema. Esses *links* são apresentados tanto para o usuário comum como para o administrador do sistema. A Figura 4.6 apresenta a tela inicial e as opções do sistema:



Figura 4.6: Página inicial da aplicação

As principais funcionalidades do sistema estão descritas brevemente a seguir:

- **administração do sistema:** acesso somente ao administrador do sistema para gerir-lo;
- **coleta de metadados:** lista opções de coleta de metadados disponíveis para consulta;
- **sistema de busca:** interface para busca de metadados. As opções disponíveis são: listar repositórios, recuperar por código, recuperar por título e consultar banco de dados;

As opções como principal, objetivo, motivação, proposta do projeto, referências e sobre o sistema são funcionalidades que complementam o aplicativo.

- **principal:** apresenta uma breve introdução e a finalidade do sistema;
- **objetivo:** lista os objetivos do sistema;
- **motivação:** apresenta algumas motivações para o desenvolvimento do protótipo;
- **proposta do projeto:** apresenta a proposta original do projeto e o desenvolvimento do protótipo;
- **referências:** lista sugestões de referências utilizadas para implementação do protótipo;
- **sobre o sistema:** traz informações sobre o sistema, bem como autores e instituições envolvidas.

A página da interface OAI para coleta de metadados é um local em comum tanto para os administradores como para os usuários comuns. Nessa página os usuários

poderão consultar os metadados de bancos de dados e das bibliotecas digitais cadastrados no sistema. A Figura 4.7 mostra as opções disponíveis para coleta de metadados:

The screenshot shows the 'Interface OAI para Coleta de Metadados' web application. At the top, there are logos for 'Repositório Digital', 'Sistema de Busca', 'UFG', and 'INSTITUTO DE INFORMÁTICA UFG'. Below the logos, the title 'Interface OAI para Coleta de Metadados' is centered. On the left side, there is a sidebar with the heading 'OPÇÕES DE PESQUISA' and several menu items: 'Informação do repositório', 'Padrão de metadado implementado pelo repositório', 'Classificação dos registros do repositório', 'Identificadores de registros do repositório', 'Localiza por título ou palavras-chave ou por assunto', 'Registros completos do repositório', and 'Repositório específico'. The main content area is titled 'Parâmetros de Pesquisa' and contains a search form. The form has two columns: 'Identificação do repositório' and 'Descrição da requisição'. The 'Identificação do repositório' column contains fields for 'Código', 'Período Inicial' (with a default value 'AAAA-MM-DD'), and 'Período Final' (with a default value 'AAAA-MM-DD'). Below these fields is a note: 'Obs.: Para recuperar todos os registros, deixe os campos em branco.' and a 'Realizar Coleta' button. The 'Descrição da requisição' column contains the text: 'Identificação do repositório (Identify): Traz as principais informações do repositório como nome, identificador, e-mail do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc.' At the bottom of the page, there is a footer with the text: 'Universidade Federal de Goiás Instituto de Informática 2010-2011 Todos os direitos reservados'.

Figura 4.7: Interfaces de coleta de metadados

As opções disponíveis da interface de coleta de metadados são:

- **informação do repositório** - identificação do repositório (*Identify*): expõe as principais informações do repositório como nome, identificador, e-mail do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc.;
- **padrão de metadado implementado pelo repositório** - formato de metadados implementados pelo sistema do repositório (*ListMetadataFormats*): lista os formatos de metadados implementados pelo sistema do repositório; é obrigatória a implementação de pelo menos o formato *Dublin Core*;
- **classificação dos registros do repositório** - classificação dos registros (*ListSets*): apresenta a árvore de assuntos que classificam os documentos no repositório ou outro conjunto de classificação;
- **identificadores de registros do repositório** - identificação de registros do repositório (*ListIdentifier*): lista todos os identificadores de registros do repositório;
- **localiza por título ou palavras-chave ou por assunto** - traz as principais informações do repositório como nome, identificador, e-mail do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc. Todas elas localizadas por título, palavras-chave ou assunto;
- **registros completos do repositório** - lista registros (*ListRecords*): lista todos os registros do repositório;
- **repositório específico** - (*GetRecord*): dado um identificador (código do registro), lista o registro correspondente.

Para as opções de consultas descritas anteriormente, existem parâmetros de pesquisa que são importantes para a orientação da busca dos metadados:

- **informação do repositório:** os parâmetros de pesquisa utilizados nesta opção são o código e os períodos inicial e final. O código é a identificação do recurso na tabela, e os parâmetros, período inicial e final, referem-se a data de publicação do recurso que se deseja recuperar, devem estar no formato AAAA-MM-DD - AAAA é o ano, MM o mês e o DD o dia. Os metadados poderão ser recuperados somente com o preenchimento do código ou dos períodos, ou do código e dos períodos informados. Para recuperar todos os registros, os parâmetros devem ser deixados em branco;
- **padrão de metadado implementado pelo repositório:** os parâmetros de pesquisa utilizados nesta opção são os mesmos descritos na consulta por informação do repositório;
- **classificação dos registros do repositório** - os parâmetros de pesquisa utilizados nesta opção são o código e a classificação do registro. O código é a identificação do recurso na tabela e a classificação pode ser um SGBD ou uma biblioteca digital. Os metadados poderão ser recuperados com o preenchimento do código, ou da classificação, ou ambos. Para recuperar todos os registros, os parâmetros devem ser deixados em branco;
- **identificação de registros do repositório** - os parâmetros de pesquisa utilizados nesta opção são todos parâmetros apresentados na consulta por informação do repositório, acrescido da classificação descrita no item anterior. Os metadados poderão ser recuperados com o preenchimento individual dos parâmetros ou por meio do agrupamento desses parâmetros, o qual recupera um registro específico. Para recuperar todos os registros, os parâmetros devem ser deixados em branco;
- **localização por título ou palavras-chave ou por assunto** - os parâmetros de pesquisa utilizados nesta opção são o título, as palavras-chave ou o assunto. Esses parâmetros devem ser utilizados isoladamente, mas, no caso de preenchimento simultâneo desses parâmetros, o sistema pesquisará somente por título. Para recuperar todos os registros, os parâmetros devem ser deixados em branco;
- **registros completos do repositório** - não há parâmetros de pesquisa. Todos os registros do repositório são recuperados;
- **repositório específico** - o parâmetro de pesquisa utilizado nesta opção é o código (identificação do recurso na tabela).

A opção que mostra os registros completos do repositório traz uma funcionalidade adicional que permite consultar o banco de dados. Essa funcionalidade é destacada na Figura 4.8. O número 1 destacado na tela é a opção que direciona o usuário a realizar pesquisa no banco de dados. O número 2 indica que o banco de dados está disponível para

relatórios que podem ser consultados. O cenário ideal para os usuários da *web* seria realizar uma pesquisa a banco de dados utilizando linguagem natural, mas neste trabalho foram disponibilizados alguns relatórios que podem ser consultados pelos usuários do sistema.

Consulta a banco de dados

Selecione um Banco de Dados para realizar a pesquisa:

- ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
- Banco de dados do CEPED referente a inscrição do IV EDIPE
- Banco de dados do projeto de pesquisa leitura na tela
- Banco de dados do Seminário de Economia da UnUCSEH/UEG
- Banco de dados da Coordenação Administrativa da UnUCSEH
- Banco de Dados Institucional da UnUCSEH
- Banco de Dados do Museu Virtual de Anápolis
- Via Litterae - Revista Eletrônica
- Banco de dados da Rede de Pesquisa REDUCATIVA
- Banco de Dados do Seminário de Contábeis da UnUCSEH
- Banco de dados da Revista Eletrônica da PrP
- Banco de dados do proj. de pesq. História e Mídia de Massa
- Banco de dados de eventos de História

Continuar a consulta a banco de dados

Figura 4.10: *Interface de consulta às informações de banco de dados*

A próxima interface disponível é o gerenciador de metadados de bancos de dados apresentado na Figura 4.11. Nesta tela o administrador poderá incluir e alterar os metadados de banco de dados, incluir e alterar metadados de bibliotecas digitais, gerar arquivo XML individual e geral, visualizar arquivo XML gerado, testar arquivos gerados no *site* do *Repository Explorer* além de coletar metadados e visualizar tabelas e registros dos bancos de dados cadastrados.

Figura 4.11: Página do administrador da aplicação

As opções apresentadas no gerenciador de metadados de bancos de dados estão descritos a seguir:

- **incluir e alterar metadados de banco de dados:** permite incluir e alterar metadados de bancos de dados fornecidos pelo administrador de banco de dados;
- **incluir e alterar metadados de bibliotecas digitais:** permite incluir e alterar metadados de bibliotecas digitais;
- **gerar arquivo no formato XML (individual e geral):** permite gerar arquivo de metadados no formato XML. Os arquivos podem ser gerados individualmente para cada banco de dados ou agrupados em um único arquivo para serem expostos para coleta de metadados;
- **visualizar arquivo XML gerado:** permite visualizar o arquivo XML dos metadados gerados para consulta na *web*;
- **testar arquivos gerados no site do Repository Explorer:** opção disponível para testar o arquivo XML gerado;
- **interface OAI para coleta de metadados:** página em que é possível realizar consultas a metadados de bancos de dados e das bibliotecas digitais cadastrados no sistema;
- **relação de bancos de dados do MySQL:** visualiza os bancos de dados configurados no servidor, bem como nomes de campos e conteúdos das tabelas.

A seguir cada uma das opções tem as funcionalidades apresentadas com detalhes.

- **Incluir e alterar metadados de banco de dados**

A Figura 4.12 apresenta a interface de inclusão de metadados no sistema. Esta interface é utilizada pelo administrador do sistema para incluir os metadados de banco de dados conforme os elementos definidos na Tabela 4.2.

[Acesso ao Menu de Opções](#)

Inclusão de Metadados

URL ou endereço do provedor do BD:

URL ou endereço do sítio vinculado ao banco de dados:

E-mail do administrador do BD:

Elementos do Dublin Core

URL do sítio onde o aplicativo está hospedado:

Disponível para consulta externa? Sim Não

Título do BD:

Nome do autor ou entidade responsável pelo recurso:

Palavras-chave do conteúdo do BD:

Resumo do Banco de Dados:

Nome da pessoa/ instituição responsável por qualquer contribuição para o conteúdo do recurso divulgado:

Nome da entidade/ instituição responsável por tornar o recurso acessível:

Data de Publicação: (AAAA-MM-DD)

Tipo do Banco de Dados: Database - MySQL

Formato do Banco de Dados: SGBDR

Nome do banco de dados conforme criado pelo administrador do banco de dados:

Principal departamento que utiliza o Banco de Dados:

Língua utilizada para descrever o recurso: Português

Sistemas que utilizam o Banco de Dados:

Área administrativa que o BD faz cobertura (financeira, cadastro, administrativo, vendas, acadêmico etc):

Informação de direitos sobre o recurso:

Login para acesso ao BD:

Senha de acesso ao BD:

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 4.12: *Página de inclusão de metadados*

- **Incluir e alterar metadados de bibliotecas digitais**

A mesma tela apresentada na Figura 4.12 é utilizada para o cadastro das bibliotecas digitais. O que as diferenciam são os campos tipo e o formato dos recursos que são configurados para biblioteca digital.

- **Gerar arquivo no formato XML (individual e geral)**

Nesta opção são gerados os arquivos XML em que são descritos os metadados de banco de dados no padrão *Dublin Core*. A extração de metadados informados pelo administrador de bancos de dados poderá ser realizada de duas formas: (1) gerar metadados de um banco de dados ou (2) gerar metadados de todos os bancos de dados que foram cadastrados.

Para ambas as formas é gerado um arquivo XML, que será armazenado em um repositório; posteriormente ele poderá ser acessado para recuperação dos metadados expostos.

- **Visualizar arquivo XML gerado**

Após o arquivo ter sido gerado, é possível visualizar o conteúdo do arquivo XML. Isso possibilita que o administrador do sistema verifique os conteúdos antes de serem publicados na *web*.

O Código [4.1](#) mostra um exemplo do arquivo XML gerado.

Código 4.1 – Arquivo de metadados gerado no formato XML

```
1  <?xml version="1.0" encoding="iso-8859-1"?>
2  <OAI-PMH>
3    <Identify>
4      <serial>1</serial>
5      <dc_title>ASTUEG-UEG</dc_title>
6      <dc_identifier>bdastueg</dc_identifier>
7      <email>betetk@gmail.com</email>
8      <dc_rights>ASTUEG</dc_rights>
9    </Identify>
10   <ListMetadataFormats>
11     <Prefix>[oai_dc]</Prefix>
12     <NameSpace>[http://www.openarchives.org/OAI/2.0/oai_dc/]</NameSpace>
13     <Schema>[http://www.openarchives.org/OAI/2.0/oai_dc.xsd]</Schema>
14   </ListMetadataFormats>
15   <ListSets>
16     <dc_type>Database - MySQL</dc_type>
17     <dc_format>SGBDR</dc_format>
18   </ListSets>
19   <ListIdentifiers>
20     <dc_identifier>bdastueg</dc_identifier>
21     <dc_date>2009-01-01</dc_date>
22     <dc_type>Database - MySQL</dc_type>
23     <dc_format>SGBDR</dc_format>
24   </ListIdentifiers>
25   <ListRecords>
26     .
27     .
28   </ListRecords>
29   <GetRecord>
30     .
31     .
32   </GetRecord>
33 </OAI-PMH>
```

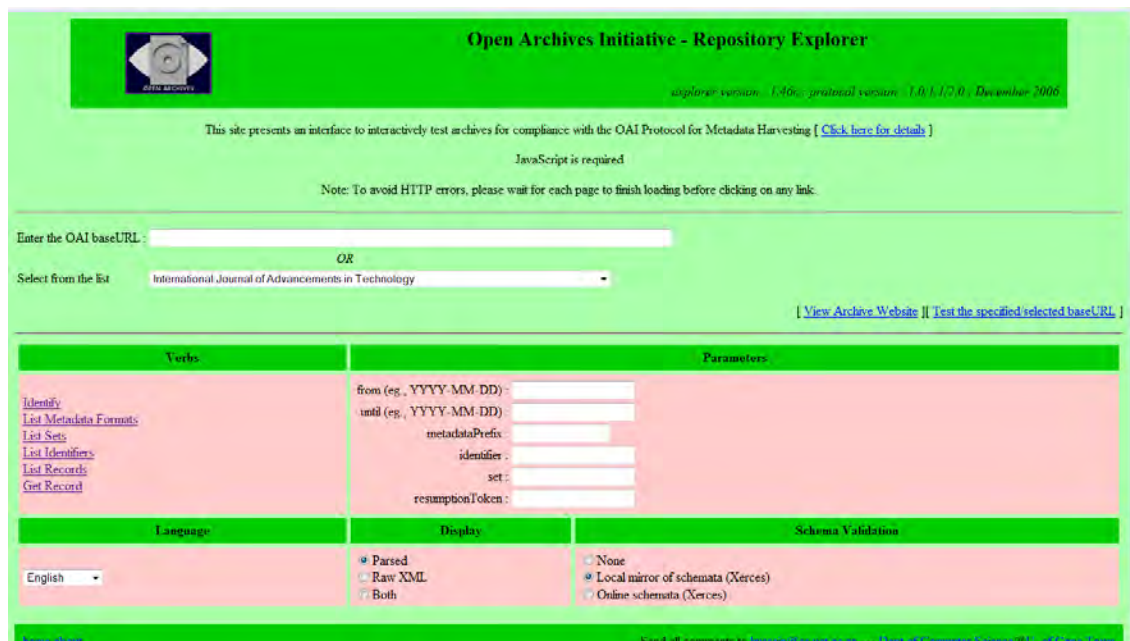


Figura 4.13: Interface do Repository Explorer [61]

- **Testar arquivos gerados no site do Repository Explorer**

Esta opção serve para testar se o arquivo XML gerado é compatível com a implementação do protocolo OAI-PMH. A interface do *Repository Explorer*[61] aparece na Figura 4.13.

- **Relação de banco de dados do MySQL**

Esta opção mostra os bancos de dados configurados no servidor, e o administrador de sistemas poderá ver os campos e os conteúdos das tabelas vinculados aos bancos de dados do sistema. O sistema foi implementado para acessar principalmente os bancos de dados *MySQL*, mas outros bancos podem ser configurados nessa opção.

A Figura 4.14 apresenta os bancos de dados configurados no servidor local. O administrador poderá visualizar os campos das tabelas do banco de dados local, bem como os conteúdos dessas tabelas, como podem ser vistos nas Figuras 4.15, 4.16 e 4.17.


Para ilustrar os bancos de dados do servidor local, utilizado para teste, encontram-se: *INFORMATION_SCHEMA*, *astueg*, *metadatabase_extractor*, *mysql*, *oai_records*, *phpmyadmin*, *tde* e *tde_submissao* (Figura 4.14).

Como exemplo de visualização de tabelas de um banco de dados foi selecionado o banco de dados *information_schema*. Como pode ser visto na Figura 4.15, é apresentada uma relação de tabelas do banco de dados. Dessa relação podem ser visualizados os campos das tabelas e os dados do banco de dados conforme aparecem nas Figuras 4.16 e 4.17.

Relacao de Banco de Dados do MySQL

Nome do Banco de Dados
information_schema
astueg
metadatabase_extractor
mysql
oai_records
phpmyadmin
tde
tde_submissao

Figura 4.14: Bancos de dados configurados no servidor local



Repositório Digital
Sistema de Busca



Relacao de Tabelas do Banco de Dados information_schema

Relação de Tabelas - information_schema	Ver dados da tabela
CHARACTER_SETS	Consulta Dados da tabela CHARACTER_SETS
COLLATIONS	Consulta Dados da tabela COLLATIONS
COLLATION_CHARACTER_SET_APPLICABILITY	Consulta Dados da tabela COLLATION_CHARACTER_SET_APPLICABILITY
COLUMNS	Consulta Dados da tabela COLUMNS
COLUMN_PRIVILEGES	Consulta Dados da tabela COLUMN_PRIVILEGES
KEY_COLUMN_USAGE	Consulta Dados da tabela KEY_COLUMN_USAGE
PROFILING	Consulta Dados da tabela PROFILING
ROUTINES	Consulta Dados da tabela ROUTINES
SCHEMATA	Consulta Dados da tabela SCHEMATA
SCHEMA_PRIVILEGES	Consulta Dados da tabela SCHEMA_PRIVILEGES
STATISTICS	Consulta Dados da tabela STATISTICS
TABLES	Consulta Dados da tabela TABLES
TABLE_CONSTRAINTS	Consulta Dados da tabela TABLE_CONSTRAINTS
TABLE_PRIVILEGES	Consulta Dados da tabela TABLE_PRIVILEGES
TRIGGERS	Consulta Dados da tabela TRIGGERS
USER_PRIVILEGES	Consulta Dados da tabela USER_PRIVILEGES
VIEWS	Consulta Dados da tabela VIEWS

Figura 4.15: Exemplo: relação de tabelas do banco de dados INFORMATION_SCHEMA



Relação de campos da Tabela TABLES

Relação de Campos - TABLES
TABLE_CATALOG
TABLE_SCHEMA
TABLE_NAME
TABLE_TYPE
ENGINE
VERSION
ROW_FORMAT
TABLE_ROWS
AVG_ROW_LENGTH
DATA_LENGTH
MAX_DATA_LENGTH
INDEX_LENGTH
DATA_FREE
AUTO_INCREMENT
CREATE_TIME
UPDATE_TIME
CHECK_TIME
TABLE_COLLATION
CHECKSUM

Figura 4.16: Exemplo: relação de campos da tabela TABLES



Dados da Tabela COLLATIONS

Campo 1	Campo 2	Campo 3	Campo 4	Campo 5
big5_chinese_ci	big5	1	Yes	Yes
big5_bin	big5	94		Yes
dec8_swedish_ci	dec8	3	Yes	
dec8_bin	dec8	69		
cp850_general_ci	cp850	4	Yes	
cp850_bin	cp850	90		
hp8_english_ci	hp8	6	Yes	
hp8_bin	hp8	72		
koi8r_general_ci	koi8r	7	Yes	
koi8r_bin	koi8r	74		
latin1_german1_ci	latin1	5		Yes
latin1_swedish_ci	latin1	8	Yes	Yes
latin1_danish_ci	latin1	13		Yes
latin1_german2_ci	latin1	31		Yes
latin1_bin	latin1	87		Yes

Figura 4.17: Exemplo: dados da tabela COLLATIONS

O capítulo seguinte apresenta os resultados obtidos durante a fase de experimentação do protótipo.

Testes/ Exemplos

O inevitável preço da confiabilidade é a simplicidade. C.A.R. Hoare.

Este capítulo apresenta os resultados obtidos das pesquisas realizadas no sistema por meio da exposição de exemplos, destacando a interoperabilidade de bancos de dados e de bibliotecas digitais. Antes da apresentação desses resultados, a Seção 5.1 evidencia a configuração do ambiente de testes.

5.1 Configuração do Ambiente de Teste

Nesta seção encontram-se as configurações básicas necessárias para a execução do teste. Para viabilizar a realização dos experimentos, uma base de dados, denominada METADATABASE_EXTRACTOR, foi criada. A base é composta de três tabelas: a primeira, e mais importante delas, registra os metadados dos bancos de dados e das bibliotecas digitais (metadbx_tb_oai_records - Código 5.1) registradas no sistema (tabela de metadados para exposição - TME). Outra armazena os dados do administrador do sistema (metadbx_cad_admdatabase - Código 5.2). A terceira armazena os relatórios disponíveis para consulta a banco de dados (metadbx_relatorios_bancodados - Código 5.3).

O teste foi realizado com quinze bancos de dados de uma universidade e duas bibliotecas digitais aleatórias cadastradas na tabela tb_oai_records definido na estrutura apresentada na Tabela 4.2.

Os detalhes dos bancos de dados e das bibliotecas digitais utilizados para teste estão descritos nas Tabelas 5.1 e 5.2 a seguir:

Tabela 5.1: Descrição dos bancos de dados utilizados para teste

Descrição	Conteúdo
Ordem	01
Título do banco de dados	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
URL da página	http://www.astueg.org.br/
Autor	ASTUEG-UEG

Resumo	Banco de dados dos associados da Associação dos Servidores Técnicos Administrativos Efetivos da UEG (Astueg).
Ordem	02
Título do banco de dados	Banco de dados do CEPED referente à inscrição do IV EDIPE
URL da página	http://www.ceped.ueg.br
Autor	Elisabete Tomomi Kowata
Resumo	Banco de dados do CEPED referente à inscrição do IV EDIPE.
Ordem	03
Título do banco de dados	Banco de dados do Centro de Documentação da UEG/UnUCSEH
URL da página	http://www.cedoc.historia.ueg.br
Autor	Elisabete Tomomi Kowata
Resumo	CEDOC/UnUCSEH Banco de dados do Centro de Documentação da UEG/UnUCSEH mantém informações e materiais históricos da instituição, da cidade de Anápolis e região de Goiás. Eles servem de material de apoio para pesquisas dos alunos da instituição.
Ordem	04
Título do banco de dados	Banco de dados do projeto de pesquisa <i>Leitura na Tela</i>
URL da página	http://www.leituranatela.ueg.br
Autor	Grupo de Pesquisa - <i>Leitura na Tela</i> /UEG
Resumo	O projeto <i>Leitura na tela</i> tem sua origem na intenção da Universidade Estadual de Goiás (UEG), juntamente com o Centro Universitário de Anápolis (UniEvangélica - UniE), de fazer um trabalho ampliado de pesquisa e de intervenção nas escolas de educação básica no Estado de Goiás. A pesquisa coordenada pela prof. Dra. Mirza Seabra Toschi evidencia um estudo do uso do laboratório de informática das escolas estaduais de Goiás e foi realizada nos municípios em que a UEG possuía curso de Pedagogia.
Ordem	05
Título do banco de dados	Banco de dados do Seminário de Economia da UnUCSEH/UEG
URL da página	http://www.seminariodeeconomia.ueg.br
Autor	UnUCSEH
Resumo	O banco de dados do Seminário de Economia da UnUCSEH/UEG mantém informação de cadastro dos participantes do evento. O Seminário teve como tema no ano de 2010: Economia Goiana Contemporânea Perspectivas e Desafios do Século XXI, que aconteceu entre os dias 23 a 27 de agosto de 2010.
Ordem	06
Título do banco de dados	Banco de dados da Coordenação Administrativa da UnUCSEH
URL da página	http://www.admunucseh.ueg.br
Autor	UnUCSEH

Resumo	O banco de dados da Coordenação Administrativa da UnUCSEH é utilizado por diversas aplicações, que não necessitam de banco de dados específico e dedicado. Portanto, é utilizado por coordenações da unidade que necessitam consultar informações rápidas e pontuais.
Ordem	07
Título do banco de dados	Banco de Dados Institucional da UnUCSEH
URL da página	http://www.unucseh.ueg.br
Autor	UnUCSEH
Resumo	O banco de dados da UnUCSEH mantém informações gerais da Unidade, desde cursos de graduação, centros e núcleos até secretarias e coordenações.
Ordem	08
Título do banco de dados	Banco de Dados do CieAA
URL da página	http://www.cieaa.ueg.br
Autor	CieAA - Centro Interdisciplinar de Estudos África-América
Resumo	O banco de dados do CieAA mantém informações do Centro com o objetivo de divulgar ideias, textos, contatos, cursos e eventos relacionados às temáticas do eixo Áfricas-Américas.
Ordem	09
Título do banco de dados	Banco de Dados do Museu Virtual de Anápolis
URL da página	http://www.museuvirtual.ueg.br
Autor	UnUCSEH
Resumo	O Museu Virtual da Educação em Anápolis é resultado de um projeto de pesquisa de professoras, alunos e técnicos da Unidade Universitária de Ciências Socioeconômicas e Humanas da Universidade Estadual de Goiás. Foram mais de dois anos de trabalho ininterrupto, apresentado no <i>site</i> que disponibiliza imagens de objetos, fotos, documentos, vídeos, textos e documentos que contam um pouco da história da educação de Anápolis, cidade que em 2011 fez 104 anos. O Museu Virtual pode ser usado por pesquisadores para ter acesso a aspectos da educação na cidade que poucos conhecem. A íntegra das entrevistas do <i>link</i> Memória Viva está à disposição dos que desejarem conhecer mais sobre as opiniões de professores conhecidos e anônimos que fizeram a educação escolar na cidade. Além disso, há disponível inúmeros documentos e fotos, livres de cupins e traças. O Museu pode ser ainda um deleite para os que desejam fazer um passeio virtual na rica e controversa história da educação básica e superior anapolina.
Ordem	10
Título do banco de dados	Via Litterae - Revista Eletrônica
URL da página	http://www.unucseh.ueg.br/vialitterae
Autor	Marco Antônio Rosa Machado (UnUCSEH/UEG)

Resumo	Via Litterae é uma revista eletrônica semestral mantida pelo Curso de Letras da Unidade Universitária de Ciências Socioeconômicas e Humanas da Universidade Estadual de Goiás (UEG). Publica estudos de caráter teórico e aplicado nas áreas de Linguística, Estudos Literários e demais áreas do conhecimento, cujo objeto de estudo é a linguagem em suas diversas manifestações. A missão da revista é ser veículo de divulgação de trabalhos relevantes, produzidos por professores e pesquisadores de instituições nacionais e estrangeiras. Seu objetivo é fomentar o diálogo acadêmico de alto nível entre pesquisadores das diversas áreas do saber que têm a linguagem como objeto de estudo.
Ordem	11
Título do banco de dados	Banco de dados da Rede de Pesquisa REDUCATIVA
URL da página	http://www.reducativa.ueg.br
Autor	UnUCSEH
Resumo	A formação da Rede de Pesquisa REDUCATIVA tem como objetivo intervir no uso pedagógico da <i>internet</i> . Visa não apenas ao uso das tecnologias de informação, mas, investi na formação de competências pedagógicas para o uso das novas mídias de forma que os capacitados operem com criatividade e alguma fluência pedagógica as Tecnologias de Informação e Comunicação (TIC) no ambiente da escola. O objetivo da Rede é capacitar docentes para a compreensão das características da leitura e da escrita hipermediáticas, a fim de que se faça um melhor uso pedagógico da virtualidade nos laboratórios de informática das escolas estaduais de educação básica de Goiás. É propósito ainda ampliar e consolidar a Rede Goiana de Pesquisa Educativa (REDUCATIVA).
Ordem	12
Título do banco de dados	Banco de Dados do Seminário de Contábeis da UnUCSEH
URL da página	http://www.seminariodecontabeis.ueg.br
Autor	UnUCSEH
Resumo	O banco de dados do Seminário de Contábeis da UnUCSEH mantém informações sobre o seminário e os participantes do evento.
Ordem	13
Título do banco de dados	Banco de dados da Revista Eletrônica da PrP
URL da página	http://www.prp.ueg.br
Autor	PrP/UEG
Resumo	O banco de dados mantém informações das Revistas Eletrônicas administradas pela PrP/UEG. As Revistas Eletrônicas da UEG utilizam o sistema <i>Open Journal System</i> (OJS), padrão internacional para publicação de revistas científicas.
Ordem	14
Título do banco de dados	Banco de dados do proj. de pesq. História e Mídia de Massa
URL da página	http://www.historiaemidiademassa.ueg.br

Autor	UnUCSEH
Resumo	O projeto <i>História e Mídias de Massa</i> pretende pesquisar a presença de elementos tradicionalmente reconhecíveis como sendo matéria-prima da análise da ciência histórica (sejam personagens, fatos ou conceitos) na construção da narrativa de produtos culturais da mídia da massa: filmes, livros, histórias em quadrinhos, vídeo <i>games</i> ou RPG. Objetiva discutir a natureza da utilização desses elementos a partir das seguintes indagações: esses elementos apresentam anacronismos ou juízo de valor? Mitificam ou desmistificam? Possuem potencial pedagógico? É possível aprender e ensinar História utilizando esses produtos de mídia de massa como ferramenta? Considerando a imensa oferta desses produtos no mercado, com maior ou menor qualidade, um recorte da realidade a ser estudada é estabelecido.
Ordem	15
Título do banco de dados	Banco de dados de eventos de História
URL da página	http://www.eventoshistoria.unucseh.ueg.br
Autor	UnUCSEH
Resumo	O banco de dados do eventos História da UnUCSEH/UEG mantém informações de eventos de história organizados por alunos e professores da Instituição.

As bibliotecas digitais utilizadas para teste foram:

Tabela 5.2: *Bibliotecas digitais utilizadas para teste*

Descrição	Conteúdo
Ordem	01
Título do banco de dados	Biblioteca Digital de Teses e Dissertações da USP
URL da página	http://www.teses.usp.br/cgi-bin/oai.pl
Autor	USP
Resumo	Biblioteca Digital de Teses e Dissertações da USP.
Ordem	02
Título do banco de dados	Biblioteca Digital de Teses e Dissertações da UFPR
URL da página	http://dspace.c3sl.ufpr.br/dspace-oai/request
Autor	UFPR
Resumo	Biblioteca Digital de Teses e Dissertações da UFPR.

Os conteúdos dos bancos de dados e das bibliotecas digitais foram registrados na tabela `metadbx_tb_oai_records`, e a estrutura está apresentada no Código 5.1:

Código 5.1 – Código SQL - Código SQL da tabela metadbx_tb_oai_records

```
1  --
2  -- Estrutura da tabela `metadbx_tb_oai_records`
3  --
4
5  CREATE TABLE IF NOT EXISTS `metadbx_tb_oai_records` (
6    `serial` int(11) NOT NULL AUTO_INCREMENT,
7    `provider` varchar(255) DEFAULT NULL,
8    `url` varchar(255) DEFAULT NULL,
9    `email` varchar(255) DEFAULT NULL,
10   `oai_set` varchar(255) DEFAULT NULL,
11   `dispquery` enum('false','true') NOT NULL
12   COMMENT 'Disponível para consulta externa',
13   `dc_title` varchar(255) DEFAULT NULL,
14   `dc_creator` text,
15   `dc_subject` varchar(255) DEFAULT NULL,
16   `dc_description` text,
17   `dc_contributor` varchar(255) DEFAULT NULL,
18   `dc_publisher` varchar(255) DEFAULT NULL,
19   `dc_date` date DEFAULT NULL,
20   `dc_type` varchar(255) DEFAULT NULL,
21   `dc_format` varchar(255) DEFAULT NULL,
22   `dc_identifier` varchar(255) DEFAULT NULL,
23   `dc_source` varchar(255) DEFAULT NULL,
24   `dc_language` varchar(255) DEFAULT NULL,
25   `dc_relation` varchar(255) DEFAULT NULL,
26   `dc_coverage` varchar(255) DEFAULT NULL,
27   `dc_rights` varchar(255) DEFAULT NULL,
28   `loginbd` varchar(15) NOT NULL
29   COMMENT 'Login do banco de dados',
30   `passwordbd` varchar(15) NOT NULL
31   COMMENT 'Senha de acesso ao banco de dados',
32   PRIMARY KEY (`serial`)
33 ) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1;
```

Os dados dos administradores do sistema são armazenados na estrutura apresentada no Código 5.2.

Código 5.2 – Código SQL - Código SQL da tabela metadbx_cad_admdatabase

```

1  --
2  -- Estrutura da tabela `metadbx_cad_admdatabase`
3  --
4
5  CREATE TABLE IF NOT EXISTS `metadbx_cad_admdatabase` (
6    `cod_adm` int(10) NOT NULL AUTO_INCREMENT,
7    `nome_adm` varchar(60) COLLATE utf8_unicode_ci NOT NULL
8    COMMENT 'nome do administrador de BD',
9    `login_adm` varchar(15) COLLATE utf8_unicode_ci NOT NULL,
10   `senha_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
11   `cpf_adm` varchar(14) COLLATE utf8_unicode_ci NOT NULL
12   COMMENT 'cpf do administrador de BD',
13   `data_nascimento_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
14   `sexo_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
15   `email_adm` varchar(60) COLLATE utf8_unicode_ci NOT NULL,
16   `data_cadastro_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL
17   COMMENT 'Data de cadastro',
18   `acesso_adm` int(1) NOT NULL
19   COMMENT 'Tipo de acesso ao sistema (9-administrador do sistema ou
20    1-administrador do banco de dados ou 0-usuario)',
21   PRIMARY KEY (`cod_adm`)
22 ) ENGINE=MyISAM DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci
23 COMMENT='Cadastro de Administradores de Banco de Dados' AUTO_INCREMENT=1;

```

Os dados de relatórios de bancos de dados disponíveis são armazenados na estrutura apresentada no [Código 5.3](#).

Código 5.3 – Código SQL - Código SQL da tabela metadbx_cad_relatorio_bancodados

```

1  --
2  -- Estrutura da tabela `metadbx_relatorio_bancodados`
3  --
4
5  CREATE TABLE IF NOT EXISTS `relatorios_bancodados` (
6    `cod_rel` int(10) NOT NULL auto_increment,
7    `query_rel` varchar(60) collate utf8_unicode_ci NOT NULL
8    COMMENT 'Descrição da consulta a banco de dados',
9    `sql_rel` varchar(250) collate utf8_unicode_ci NOT NULL
10   COMMENT 'linha de consulta sql a ser submetido ao banco de dados',
11   `serial_bd` int(10) NOT NULL,
12   PRIMARY KEY (`cod_rel`)
13 ) ENGINE=MyISAM DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci
14 COMMENT='Cadastro Relatórios' AUTO_INCREMENT=1 ;

```

A Seção 5.2 apresenta exemplos de consultas submetidas ao sistema e os resultados dessas consultas.

5.2 Exemplos de consultas

O sistema disponibiliza dois *links* de pesquisa: coleta de metadados e sistema de busca. A coleta de metadados possibilita a pesquisa de informações de seis formas: informação do repositório, padrão de metadado implementado pelo repositório, classificação dos registros do repositório, identificadores de registros do repositório, localização por título, palavras-chave ou assunto, registros completos do repositório e repositório específico. O sistema de busca possui quatro opções de consulta: listar repositório, recuperar pelo código, recuperar por título e consultar banco de dados. A opção de coleta de metadados é direcionada ao administrador do sistema, enquanto o sistema de busca é direcionado aos usuários em geral. A seguir são apresentadas as interfaces de busca das opções mencionadas, iniciando com a pesquisa de coleta de metadados.

5.2.1 Coleta de Metadados

A primeira interface é a de coleta de metadados que traz as principais informações do repositório. Nesta opção os parâmetros disponíveis para consulta são o código e o período inicial e final. A Figura 5.2 apresenta os resultados para os parâmetros informados como brancos, que listará todos os repositórios encontrados.

Interface OAI para Coleta de Metadados

Parâmetros de Pesquisa

:: Identificação do repositório		Descrição da requisição
Código:	<input type="text"/>	<i>Identificação do repositório (Identify):</i> Traz as principais informações do repositório como nome, identificador, e-mail do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc.
Período Inicial:	<input type="text"/> AAAA-MM-DD	
Período Final:	<input type="text"/> AAAA-MM-DD	
Obs.: Para recuperar todos os registros, deixe os campos em branco.		
<input type="button" value="Realizar Coleta"/>		

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 5.1: Interface para coleta de metadados: parâmetros de Identify

O relatório traz informações como nome, identificador, *e-mail* do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc.

The screenshot shows the 'Repositório Digital' interface with the title 'Resultado da Pesquisa - Identificação do repositório'. It lists search parameters and their corresponding values:

Nome do Repositório:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	betek@gmail.com
Direitos sobre o recurso:	ASTUEG

Figura 5.2: Resultado da coleta de metadados: Identify

A interface de coleta de metadados que diz respeito ao padrão de metadados implementado pelo repositório é apresentada na Figura 5.3. Nesta opção os parâmetros disponíveis para consulta são o código e o período inicial e final. A Figura 5.4 apresenta os resultados para os parâmetros informados como brancos. Neste relatório são listados: código de identificação, título, nome, *e-mail* do administrador, direitos sobre o recurso e formato de metadados utilizados.

The screenshot shows the 'Interface OAI para Coleta de Metadados' with a sidebar menu and a search parameter form. The sidebar menu includes options like 'Informação do repositório', 'Padrão de metadado implementado pelo repositório', 'Classificação dos registros do repositório', etc. The search parameters form is as follows:

Parâmetros de Pesquisa		Descrição da requisição
:: Padrão de metadado implementado pelo repositório		Formato de metadados implementados pelo sistema do repositório (ListMetadataFormats). Lista os formatos de metadados implementados pelo sistema do repositório; é obrigatória a implementação de pelo menos o formato Dublin Core.
Código:		
Período Inicial:	AAAA-MM-DD	
Período Final:	AAAA-MM-DD	
Obs.: Para recuperar todos os registros, deixe os campos em branco.		
<input type="button" value="Realizar Coleta"/>		

At the bottom of the interface, it states: Universidade Federal de Goiás, Instituto de Informática, 2010-2011 Todos os direitos reservados.

Figura 5.3: Interface para coleta de metadados: parâmetros de ListMetadataFormats

Resultado da Pesquisa - Padrão de metadado implementado pelo repositório

Coletando em ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG - .../bd1/database/oi/oi_dc_oaireg1.xml
 Parâmetros de pesquisa estão vazios.

Código de Identificação:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	beterk@gmail.com
Direitos sobre o recurso:	ASTUEG
Formato de Metadados utilizado:	
Prefixo:	[oi_dc]
Namespace:	[http://www.openarchives.org/OAI/2.0/oi_dc]
Schema:	[http://www.openarchives.org/OAI/2.0/oi_dc.xsd]

Coletando em Banco de dados do CEPED referente a inscrição do IV EDIPE - .../bd2/database/oi/oi_dc_oaireg2.xml
 Parâmetros de pesquisa estão vazios.

Código de Identificação:	2
Título do banco de dados:	Banco de dados do CEPED referente a inscrição do IV EDIPE
Nome do banco de dados:	dbceped
E-mail do administrador de banco de dados:	beterk@gmail.com
Direitos sobre o recurso:	CEPED - Centro de Estudos e Pesquisa em Didática
Formato de Metadados utilizado:	
Prefixo:	[oi_dc]
Namespace:	[http://www.openarchives.org/OAI/2.0/oi_dc]
Schema:	[http://www.openarchives.org/OAI/2.0/oi_dc.xsd]

Figura 5.4: Resultado da coleta de metadados: *ListMetadataFormats*

A interface de coleta que apresenta a classificação dos registros do repositório é apresentada na Figura 5.5. Nessa opção os parâmetros disponíveis para consulta são o código e a classificação do repositório (SGBDR ou biblioteca digital). A Figura 5.6 apresenta os resultados para os parâmetros informados como brancos, que listará tanto os bancos de dados como as bibliotecas digitais. Ao selecionar a classificação do repositório, listará somente o tipo selecionado. As informações apresentadas nesse relatório são o código de identificação, título, nome, *e-mail* do administrador, direitos sobre o recurso, tipo e formato do repositório.

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Interface OAI para Coleta de Metadados

OPÇÕES DE PESQUISA

- Informação do repositório
- Padrão de metadado implementado pelo repositório
- Classificação dos registros do repositório**
- Identificação dos registros do repositório
- Localiza por título ou palavras-chave ou por assunto
- Registros completos do repositório
- Repositório específico

Parâmetros de Pesquisa

Classificação dos registros do repositório	Descrição da requisição
Código: <input type="text"/>	<i>Classificação dos registros (ListSets):</i> Traz a árvore de assuntos que classificam os documentos no repositório ou outro conjunto de classificação.
Classificação: <input type="text"/>	
Obs.: Para recuperar todos os registros, deixe os campos em branco.	
<input type="button" value="Realizar Coleta"/>	

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 5.5: Interface para coleta de metadados: parâmetros de ListSets

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Resultado da Pesquisa - Classificação dos registros do repositório

Coletando em ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG - .../bd1/database/oai/dc_oaireg1.xml
Parâmetros de pesquisa estão vazios.

Código de Identificação:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	betec@igmail.com
Direitos sobre o recurso:	ASTUEG
Classificação dos repositórios	
Tipo:	Database - MySQL
Formato:	SOBDR

Coletando em Banco de dados do CEPED referente a inscrição do IV EDIPE - .../bd2/database/oai/dc_oaireg2.xml
Parâmetros de pesquisa estão vazios.

Código de Identificação:	2
Título do banco de dados:	Banco de dados do CEPED referente a inscrição do IV EDIPE
Nome do banco de dados:	@ceped
E-mail do administrador de banco de dados:	betec@igmail.com
Direitos sobre o recurso:	CEPED - Centro de Estudos e Pesquisa em Didática
Classificação dos repositórios	
Tipo:	Database - MySQL
Formato:	SOBDR

Coletando em Biblioteca Digital de Teses e Dissertações da USP - .../bd3/database/oai/dc_oaireg3.xml

Figura 5.6: Resultado da coleta de metadados: ListSets

A pesquisa que mostra todos os identificadores de registros tem como parâmetros de busca o código, o período inicial e final e a classificação do repositório. A Figura 5.8 apresenta os resultados para os parâmetros informados como brancos.

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Interface OAI para Coleta de Metadados

OPÇÕES DE PESQUISA

- Informação do repositório
- Padrão de metadado implementado pelo repositório
- Classificação dos repositórios
- Identificadores de registros do repositório**
- Palavras-chave ou por assunto
- Registros completos do repositório
- Repositório específico

Parâmetros de Pesquisa

Identificadores de registros do repositório		Descrição da requisição
Código:	<input type="text"/>	<i>Identificação de registros do repositório (ListIdentifier):</i> Lista todos os identificadores de registros do repositório.
Período Inicial:	AAAA-MM-DD	
Período Final:	AAAA-MM-DD	
Classificação:	<input type="text"/>	
Obs.: Para recuperar todos os registros, deixe os campos em branco.		
<input type="button" value="Realizar Coleta"/>		

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 5.7: Interface para coleta de metadados: parâmetros de *ListIdentifier*

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Resultado da Pesquisa - Identificadores de registros do repositório

Coletando em ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG - .../bd1/database/oai/dc_oaireg1.xml

Parâmetros de pesquisa estão vazios.

Código de identificação:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	betete@gmail.com
Direitos sobre o recurso:	ASTUEG 2010-01-01
Tipo do banco de dados:	Database - MySQL
Formato do banco de dados:	SGDR

Coletando em Banco de dados do CEPED referente a inscrição do IV EDIPE - .../bd2/database/oai/dc_oaireg2.xml

Parâmetros de pesquisa estão vazios.

Código de identificação:	2
Título do banco de dados:	Banco de dados do CEPED referente a inscrição do IV EDIPE
Nome do banco de dados:	dbceped
E-mail do administrador de banco de dados:	betete@gmail.com
Direitos sobre o recurso:	CEPED - Centro de Estudos e Pesquisa em Didática 2011-03-01
Tipo do banco de dados:	Database - MySQL
Formato do banco de dados:	SGDR

Coletando em Biblioteca Digital de Teses e Dissertações da USP - .../bd3/database/oai/dc_oaireg3.xml

Figura 5.8: Resultado da coleta de metadados: *ListIdentifier*

A interface de pesquisa mostrada na Figura 5.9 tem como parâmetros de busca o título, as palavras-chave ou o assunto. Essa pesquisa permite recuperar todos os registros que apresentam alguma ocorrência em um desses campos informados. A Figura 5.10 expõe os resultados para os parâmetros informados como brancos.

Repositório Digital
Versão Beta Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Interface OAI para Coleta de Metadados

OPÇÕES DE PESQUISA

- Informação do repositório
- Padrão de metadado implementado pelo repositório
- Classificação dos registros do repositório
- Identificadores de registros do repositório
- Localiza por título ou palavras-chave ou por assunto**
- Registros completos do repositório
- Repositório específico

Parâmetros de Pesquisa

:: Localiza por título ou palavras-chave ou por assunto

Título:	
Palavra Chave:	
Assunto:	
<input type="button" value="Realizar Coleta"/>	

Descrição da requisição

Traz as principais informações do repositório como nome, identificador, e-mail do administrador, informações sobre a propriedade intelectual dos dados contidos no repositório etc localizados por título, palavras-chave ou assunto.

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 5.9: Interface para coleta de metadados: parâmetros de pesquisa por título ou palavras-chave ou assunto

Repositório Digital
Versão Beta Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Resultado da Pesquisa - Por título ou palavras-chave ou assunto

Parâmetros de pesquisa estão vazios.

Número de registros: 21

Código de Identificação:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	betark@gmail.com
Direitos sobre o recurso:	ASTUEG
Ver detalhes >>>	
Código de Identificação:	2
Título do banco de dados:	Banco de dados do CEPED referente a migração do IVEDIPE
Nome do banco de dados:	ibcped
E-mail do administrador de banco de dados:	betark@gmail.com
Direitos sobre o recurso:	CEPED - Centro de Estudos e Pesquisa em Didática
Ver detalhes >>>	
Código de Identificação:	1
Título do banco de dados:	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados:	astueg
E-mail do administrador de banco de dados:	betark@gmail.com
Direitos sobre o recurso:	ASTUEG
Ver detalhes >>>	

Figura 5.10: Resultado da coleta de metadados: pesquisa por título ou palavras-chave ou assunto

A pesquisa que mostra todos os registros é a opção *Registros completos do repositório*; não é necessário, com isso, informar nenhum parâmetro para a busca. A Figura 5.12 apresenta os resultados dessa opção.



The screenshot shows the 'Repositório Digital' search interface. At the top, there is a logo for 'banco de dados Metadatabase Extractor' and 'Repositório Digital Sistema de Busca'. The main heading is 'Interface OAI para Coleta de Metadados'. On the left, there is a sidebar with 'OPÇÕES DE PESQUISA' and several menu items, with 'Registros completos do repositório' circled in red. The main area is titled 'Parâmetros de Pesquisa' and contains a search form with the text 'Lista todos os registros do repositório' and a 'Realizar Coleta' button. To the right, there is a 'Descrição da requisição' section with the text 'Lista registros (ListRecords): Lista todos os registros do repositório.' At the bottom, there is a footer with 'Universidade Federal de Goiás Instituto de Informática 2010-2011 Todos os direitos reservados'.

Figura 5.11: Interface para coleta de metadados: parâmetros de ListRecords



The screenshot shows the 'Resultado da Pesquisa - Lista todos os registros do repositório' section. It displays the following metadata for a record:

```

Coletando em ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG - .../bd1/database/oai/dc_oaireg1.xml
Código de Identificação: 1
Título do banco de dados: ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG
Nome do banco de dados: astueg
E-mail do administrador de banco de dados: beterk@gmail.com
Direitos sobre o recurso: ASTUEG
-- META-DADOS DO REPOSITÓRIO --
-- META-DADOS DO REPOSITÓRIO --
Previdor: .../bd1/database/oai/dc_oaireg1.xml
URL do banco de dados: http://www.astueg.org.br
URL do sistema: http://www.astueg.org.br
Banco de Dados Disponível: sim
Autor: ASTUEG UEG
Palavras-chave: Dados dos associados da Astueg
Resumo: Banco de dados dos associados da Astueg - Associação dos Servidores Técnicos Administrativos Efetivos da UEG.
Contribuidor: Inessa Machado Lima
Responsável pela publicação: Elizabete Tomoma Kovata
Data de publicação: 2010-01-01
Tipo de recurso: Database - MySQL
Formato do recurso: SOBDR
Nome do banco de dados: astueg
  
```

Figura 5.12: Resultado da coleta de metadados: ListRecords

E por fim a última pesquisa disponível ao usuário encontrada neste grupo tem como parâmetro de busca o código do registro. A Figura 5.14 apresenta o resultado dessa pesquisa.

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Interface OAI para Coleta de Metadados

OPÇÕES DE PESQUISA

- Informação do repositório
- Padrão de metadado implementado pelo repositório
- Classificação dos registros do repositório
- Identificadores de registros do repositório
- Localiza por título ou palavras-chave ou por assunto
- Registros completos do repositório
- Repositório específico**

Parâmetros de Pesquisa

:: Repositório específico		Descrição da requisição
Código:	<input type="text"/>	(GetRecord): Dado um identificador (código do registro), lista o registro correspondente.
Realizar Coleta		

Universidade Federal de Goiás
Instituto de Informática
2010-2011 Todos os direitos reservados

Figura 5.13: Interface para coleta de metadados: parâmetros de GetRecord

Repositório Digital
Versão Beta
Sistema de Busca

banco de dados
Metadatabase Extractor

UFG INSTITUTO DE INFORMÁTICA UFG

Resultado da Pesquisa - Informação de um Banco de Dados por código identificador

Parâmetro de pesquisa: código

Código de Identificação: 4
Título do banco de dados: Repositório Digital da UFRGS
Nome do banco de dados:
E-mail do administrador de banco de dados: emailadministrador@bib.com.br
Direitos sobre o recurso: UFRGS

DETAHES DO REGISTRO ::

Provedor:	http://www.lume.ufrgs.br/dspace-ou-requisi
URL do banco de dados:	
URL do sistema:	
Banco de Dados Disponível:	sim
Autor:	UFRGS
Palavras-chave:	Biblioteca digital UFRGS
Resumo:	
Contribuidor:	UFRGS
Responsável pela publicação:	UFRGS
Data de publicação:	2011-06-29
Tipo do recurso:	Biblioteca Digital
Formato do recurso:	Biblioteca Digital
Nome do banco de dados:	
Departamento que mantém o BD:	

Figura 5.14: Resultado da coleta de metadados: GetRecord

5.2.2 Sistema de Busca

Além da coleta de metadados apresentados anteriormente, o sistema disponibiliza um mecanismo de busca com as seguintes opções: listar repositórios, recuperar por código, recuperar por título e consultar banco de dados (Figura 4.9).

A opção "listar repositórios" lista todos os repositórios cadastrados no sistema. O resultado listado pelo sistema aparece na Figura 5.15

Código	Título	URL Página	Formato	Disponível para consulta
1	ASTUEG-Assoc. dos Servidores Técnicos-Adm. Efetivos da UEG	http://www.astueg.org.br/	SGBDR	sim
14	Banco de dados da Coordenação Administrativa da UnUCSEH	http://www.admunucseh.ueg.br	SGBDR	sim
19	Banco de dados da Rede de Pesquisa REDUCATIVA	http://www.reducativa.ueg.br	SGBDR	sim
21	Banco de dados da Revista Eletrônica da PrP	http://www.prp.ueg.br	SGBDR	sim
23	Banco de dados de eventos de História	http://www.eventoshistoria.unucseh.ueg.br	SGBDR	sim
11	Banco de dados do Centro de Documentação da UEG/UnUCSEH	http://www.cedoc.historia.ueg.br	SGBDR	não
2	Banco de dados do CEPED referente a inscrição do IV EDIPE	http://www.ceped.ueg.br	SGBDR	sim
16	Banco de Dados do CieAA	http://www.cieaa.ueg.br	SGBDR	não
17	Banco de Dados do Museu Virtual de Anápolis	http://www.museuvirtual.ueg.br	SGBDR	sim
22	Banco de dados do proj. de pesq. História e Mídia de Massa	http://www.historiaemidiademassa.ueg.br	SGBDR	sim
12	Banco de dados do projeto de pesquisa leitura na tela	http://www.leituranatela.ueg.br	SGBDR	sim
20	Banco de Dados do Seminário de Contábeis da UnUCSEH	http://www.seminariodecontabeis.ueg.br	SGBDR	sim
13	Banco de dados do Seminário de Economia da UnUCSEH/UEG	http://www.seminariodeeconomia.ueg.br	SGBDR	sim
15	Banco de Dados Institucional da UnUCSEH	http://www.unucseh.ueg.br	SGBDR	sim
6	Biblioteca Digital de Teses e Dissertações da UFPR	http://dspace.c3sl.ufpr.br/dspace-oai/request	Biblioteca Digital	sim
7	Biblioteca Digital de Teses e Dissertações da UFRN	http://bdt.d.bc2m.ufrr.br/tesesimplificado/tdc_oai/oai2.php	Biblioteca Digital	não
3	Biblioteca Digital de Teses e Dissertações da USP	http://www.theses.usp.br/cgi-bin/oai.pl	Biblioteca Digital	sim
5	Biblioteca Digital Jurídica do Superior Tribunal de Justiça	http://bdjur.stj.gov.br/dspace-oai/request	Biblioteca Digital	não
8	Repositório Ação Educativa	http://www.bdae.org.br/dspace-oai/request	Biblioteca Digital	não
9	Repositório Diálogo Científico	http://bvc.cgu.gov.br/dspace-oai/request	Biblioteca Digital	não
4	Repositório Digital da UFRGS	http://www.lume.ufrgs.br/dspace-oai/request	Biblioteca Digital	não
10	Repositório Institucional da UnB	http://bvc.cgu.gov.br/dspace-oai/request	Biblioteca Digital	não
18	Via Litterae - Revista Eletrônica	http://www.unucseh.ueg.br/vialitterae	SGBDR	sim

Figura 5.15: Resultado de busca: listar repositórios

A consulta a banco de dados será possível desde que o *status* para pesquisa a banco de dados esteja preenchido com "sim". Na Figura 4.10 apresentada no Capítulo 4 aparece uma relação de bancos de dados em que o *status* aparece como disponível para consulta. Ao selecionar um banco de dados da lista, será apresentada uma listagem de relatórios que o usuário poderá selecionar para consulta.

Isso demonstra que também é possível disponibilizar as descrições de bancos de dados possibilitando que as informações sejam consultadas como são feitas com as bibliotecas digitais.

Conclusão

*A tecnologia ensinou uma lição à Humanidade: nada é impossível.
Lewis Mumford, filósofo americano, 1895-1990.*

Promover interoperabilidade entre fontes de informações heterogêneas é uma preocupação antiga de pesquisadores de diversas áreas, incluindo banco de dados. Porém, com o surgimento da *internet* a dimensão da real necessidade tornou-se evidente, não apenas dos dados disponíveis na rede, mas também no âmbito das organizações. No que se refere a este trabalho, o objetivo principal é permitir que bancos de dados relacionais possam ser consultados por meio da *internet*, tornando-os interoperáveis com outras fontes de informações, como por exemplo as bibliotecas digitais. Isso permitirá que as informações, antes restritas às organizações, possam ser compartilhadas na grande rede mundial.

Mais especificamente, no caso de banco de dados relacionais, promover a interoperabilidade com demais fontes de informações exige que certos aspectos técnicos e administrativos sejam resolvidos, tais quais: a representação e a recuperação de metadados desses repositórios, a exposição das informações, o acesso a bancos de dados e a criação de mecanismos automáticos de busca a bancos de dados na *web* dentre outros. O trabalho mostrou que nos bancos de dados relacionais, informações como a localização, o proprietário, a descrição do conteúdo etc. não são, em geral, armazenadas no catálogo de um sistema gerenciador de banco de dados, e para obtê-las é necessário recorrer aos conhecimentos do administrador do banco de dados.

Para superar as limitações dos catálogos desses bancos de dados, o trabalho apresenta uma proposta de extensão em que essas informações são armazenadas, possibilitando que elas possam ser disponibilizadas na *web*. A proposta de criação de uma extensão do catálogo de bancos de dados relacionais até sua publicação é descrita em três etapas: descrição dos metadados de um banco de dados por meio do padrão *Dublin Core*, extração dos metadados que descrevem o conteúdo do banco de dados e utilização do protocolo OAI-PMH para exposição dos metadados extraídos.

Na etapa que descreve o conteúdo de um banco de dados por meio do padrão *Dublin Core*, foram estudados os elementos deste padrão utilizados para bibliotecas digitais e, desse estudo, foram elencados os elementos que tiveram o maior número de ocorrências, tanto para os elementos simplificados como para os elementos completos. Nesse levantamento observou-se que as bibliotecas digitais pesquisadas utilizam de forma diversa alguns elementos do padrão. Portanto, há uma variação do significado dos elementos do padrão *Dublin Core* adotados pelas bibliotecas digitais. O estudo ofereceu subsídio para definir os elementos do *Dublin Core* para descrever o banco de dados e acrescentar outros elementos adicionais para complementar a descrição.

O resultado do trabalho foi o desenvolvimento de uma ferramenta para representação e recuperação de metadados de banco de dados relacionais por meio da qual será possível permitir consultas, possivelmente em linguagem natural, às informações armazenadas nesses repositórios, viabilizando também a interoperabilidade entre fontes de dados heterogêneas. Pode-se demonstrar com o trabalho que as informações dos bancos de dados podem ser disponibilizadas na rede, sem contudo comprometer a segurança dos dados armazenados nos bancos de dados, pois a autorização para disponibilizá-los continua ainda sob a responsabilidade do administrador do banco de dados.

Das motivações relacionadas no início do trabalho, pode-se perceber grande contribuição com a *hidden web*, bibliotecas digitais e administração de dados.

Portanto, a importância do trabalho está na administração e gerenciamento dos bancos de dados de uma organização, na formação de uma base de dados geral de informações de bancos de dados, na ampliação das fontes de pesquisas, possibilidade dos usuários realizarem consultas a esses dados por meio de mecanismos de buscas existentes, além de contribuir com a recuperação de informação que se encontra escondida na *web*.

Os próximos passos da pesquisa incluem, entre outros objetivos, produzir uma versão do sistema em que seja possível a consulta do conteúdo de um banco de dados e o uso de ontologias como mecanismo auxiliar na interpretação de consultas por área de conhecimento.

Dessa forma, a Seção 6.1 sintetiza as principais contribuições deste trabalho. A Seção 6.2 descreve possíveis desdobramentos e trabalhos futuros. Finalmente a Seção 6.3 relaciona os relatórios técnicos produzidos durante o desenvolvimento desta dissertação.

6.1 Contribuições

Nesta seção apresenta-se um resumo das principais contribuições identificadas durante o desenvolvimento desta dissertação.

- *metadados extraídos do banco de dados e sua correspondência no padrão Dublin Core*: A primeira contribuição que este trabalho aponta é o estudo de metadados

extraídos do banco de dados e sua correspondência no padrão *Dublin Core*. O padrão é aberto e pode ser utilizado conforme a necessidade levantada pelo implementador. A pesquisa envolveu o estudo dos elementos *Dublin Core* e a forma que seriam utilizados para descrever a extensão do catálogo de bancos de dados.

- *desenvolvimento de uma ferramenta*: A segunda contribuição relevante é o desenvolvimento e a implementação de um método para extração e exposição de metadados por meio do protocolo OAI-PMH de um conjunto de banco de dados relacionais.
- *extensão do catálogo de banco de dados para pesquisa*: Disponibilizar um catálogo de banco de dados para o acesso da comunidade de usuários e oferecê-lo como fonte de pesquisa e referência para os estudos voltados a várias áreas de conhecimento.
- *interoperabilidade de banco de dados com outros repositórios*: Auxiliar pesquisas na área de interoperabilidade de bancos de dados com outros repositórios da *web* facilitando a experimentação de propostas para recuperação de informações de banco de dados relacionais.
- *hidden web*: O trabalho vem contribuir com a *hidden web*, no sentido de permitir que as informações armazenadas em bancos de dados possam ser recuperadas e compartilhadas na *web*. Esse espaço isolado considerado "ilha de informação" poderá ser integrado na grande rede mundial e permitir que mais usuários possam ter acesso ao seu conteúdo.
- *bibliotecas digitais*: A contribuição para as bibliotecas digitais está na ampliação das fontes de pesquisa. As pesquisas em bibliotecas digitais abarcam fontes como artigos, periódicos, dissertações, imagens e músicas. Com este trabalho, as bibliotecas digitais podem estender a pesquisa também a bancos de dados.
- *administração dos bancos de dados existentes em uma organização*: A contribuição imediatamente visível está na administração dos bancos de dados existentes em uma organização. O mecanismo proposto permitirá que os administradores de bancos de dados possam reunir os bancos de dados dispersos, catalogá-los e compartilhá-los dentro da organização. Ainda na administração dos bancos de dados, outra contribuição está na formação de uma base de dados geral de informações sobre bancos de dados. O proprietário do banco de dados que desejar compartilhar com

demais usuários da *web*, bastará informar a descrição do banco de dados e oferecer mecanismo de pesquisa a essas informações. Esse passo permitirá que o banco de dados possa ser indexado e os dados compartilhados com demais usuários da grande rede.

6.2 Trabalhos Futuros

Durante o desenvolvimento da pesquisa, vislumbrou-se alguns aspectos que podem ser objetos de futuras pesquisas, no sentido de estender os estudos aqui realizados.

- **expandir a pesquisa para além dos bancos de dados relacionais.** Neste trabalho, os bancos de dados utilizados seguem o paradigma relacional, principalmente devido à grande parte dos dados existentes nas organizações ser armazenada nesse tipo de repositório. Os estudos poderiam ser estendidos para outros tipos de bancos de dados, além dos relacionais, possibilitando que outros bancos sejam inseridos no sistema.
- **expandir o estudo para consulta contextualizada por meio do uso de ontologias.** Com o uso de ontologias será possível expandir a consulta para recuperar o maior número possível de informações, pois ele unifica termos, conceitos, categorias e relações de um mesmo domínio.
- **expandir o estudo para consulta em banco de dados por meio de linguagem natural.** Os mecanismos de buscas como Google e Yahoo aproximam a linguagem do ser humano para realizar consultas, porém ainda não foi possível chegar ao entendimento que o ser humano possui, transferindo o trabalho de seleção das informações recuperadas para o pesquisador. Apesar da limitação, os mecanismos de busca têm desempenhado satisfatoriamente as consultas submetidas a ele. Esse desempenho visto em mecanismos de busca e utilizado para o banco de dados possibilitaria que usuários comuns possam pesquisar em banco de dados sem necessidade de conhecimento de linguagem específica para consulta a esse repositório.
- **estudar/ trabalhar buscas integradas.** Este trabalho apresentou uma ideia de pesquisa a metadados de bancos de dados para disponibilizá-los na grande rede. Por enquanto o sistema realiza pesquisas aos banco de dados cadastrados e de metadados coletados de bibliotecas digitais. Um estudo promissor é integrar a pesquisa local com a pesquisa na grande rede, integrando dessa forma a busca.

- **uso de ontologias/ taxonomias nos metadados.** A utilização de ontologias e definição de taxonomias nos metadados de bancos de dados relacionais poderá ser usado para armazenar e manipular conhecimento para a concepção de inferências, tomadas de decisões ou apenas para responder a perguntas.
- **uso de serviços *web*/ agentes inteligentes.** A coleta de dados poderá ser facilitada por meio do uso de serviços *web* e/ou agentes inteligentes que monitorem a lista de endereços determinada pelo usuário. A busca e seleção por repositórios é uma tarefa manual e exaustiva, que poderiam ser executados por agentes inteligentes.
- **geração automática de metadados de bancos de dados relacionais.** Gerar automaticamente os metadados de banco de dados relacionais é uma atividade que exige conhecimento das estruturas de armazenamento e da sintaxe da linguagem de acesso, além disso, informações como descrição do conteúdo do banco de dados, a localização, o proprietário etc. não são, em geral, armazenadas no catálogo de um sistema gerenciador de banco de dados. Para obter essas informações é necessário recorrer aos conhecimentos do administrador desse banco. Com a geração automática de metadados, a dependência total do administrador pode ser minimizada, mas não eliminada.

6.3 Produção Bibliográfica

Durante o desenvolvimento deste trabalho foram produzidos dois relatórios técnicos com os títulos "Consulta a banco de dados guiadas por ontologias"[38] e "Metadados no contexto de banco de dados relacionais: extração e exposição com o protocolo OAI-PMH"[39]. Além dos relatórios técnicos a pesquisa foi apresentada em dois eventos da Universidade Federal de Goiás: Conpeex 2010 e no Seminário II do Instituto de Informática.

Referências Bibliográficas

- [1] ANSI/PMI. **Um guia do conjunto de conhecimentos em gerenciamento de projetos: guia PMBOK**. Project Management Institute, Four Campus Boulevard, Newtown Square, PA, United States, 2004.
- [2] ARMS, W. **A spectrum of interoperability: the site for science prototype for the nsdl**. *Dlib Magazine*, 8 (1), 2002.
- [3] BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. ACM Press, United States, 1999.
- [4] BONIFACIO, A. S.; HEUSER, C. A. **Metadados semânticos para buscas em bibliotecas digitais**. Disponível em: <<http://www.uel.br/pessoal/ailton/Trabalhos/SemanaAcad-Ailton.html>>. Acesso em: 31 maio 2010, 2010.
- [5] BRASIL, C. O. **Open archives initiative**. Disponível em <<http://clubeoai.iv.org.br/portal/oai>>. Acesso em: 21 set. 2010, 2010.
- [6] CAFÉ, L.; FACHIN, G. R. B. **Provedores de dados, provedores de serviços e periódicos em ciência da informação, biblioteconomia e áreas afins**. *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 1º sem. 2007*, 2007.
- [7] CAPLAN, P. **International metadata initiatives: lessons in bibliographic control**. BIOCENTENNIAL CONFERENCE ON BIBLIOGRAPHIC CONTROL IN THE NEW MILLENNIUM. Washington: Library of Congress, 23 jan. 2001. Disponível em: <<http://lcweb.loc.gov/catdir/bibcontrol/caplan.html>>. Acesso em: 23 set. 2010, 2010.
- [8] CONTESSA, D. F. **Um serviço de geração de metadados compatível com o padrão oai para o sistema jems, dissertação de mestrado em ciência da computação**. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.
- [9] CONTESSA, D. F.; LAZZAROTTO, D.; DE OLIVEIRA, J. P. M. **Um provedor de dados para bibliotecas digitais compatível com o padrão oai**. *In: 1º Workshop em Bibliotecas Digitais (WDL2005), SBBD/SBES, Uberlândia, 2005*.

- [10] DA SILVA FAGUNDES, A. **Projeto e implementação de um banco de dados de metadados para o sistema de informação de biodiversidade do estado de são paulo**. Master's thesis, Dissertação de Mestrado, Universidade Estadual de Campinas, Campinas, São Paulo, 1999.
- [11] DATE, C. J. **Guia para o DB2**. Campus, Rio de Janeiro, 1989.
- [12] DATE, C. J. **Uma introdução a sistemas de banco de dados**. Campus, Rio de Janeiro, 2003.
- [13] DE MENEZES CARDOSO JUNIOR, M. J. **Clio-i: Interoperabilidade entre repositórios digitais utilizando o protocolo oai-pmh**. Master's thesis, Universidade Federal de Pernambuco, Centro de Informática, Pós-Graduação em Ciência da Computação. Dissertação de Mestrado, Recife, 2007.
- [14] DE OLIVEIRA, R. R. **Recuperação contextualizada de documentos integrados pelo protocolo oai-pmh**. Master's thesis, Programa de Pós-Graduação em Ciência da Computação, Instituto de Informática, INF-UFG, 2010.
- [15] DE OLIVEIRA, R. R.; CAMILO, C. O.; DE CARVALHO, C. L.; DA SILVA, J. C. **Recuperação contextualizada de documentos em bibliotecas digitais integradas**. *In: 8th International Information and Telecommunication Technologies Symposium, Florianópolis/SC. Proceedings of the 8th International Information and Telecommunication Technologies Symposium, 2009.*
- [16] DE OLIVEIRA, R. R.; DE CARVALHO, C. L. **Implementação de interoperabilidade entre repositórios digitais por meio do protocolo oai-pmh**. Internal report, Instituto de Informática da Universidade Federal de Goiás, março 2009.
- [17] DE OLIVEIRA, V. S. **Buscando interoperabilidade entre diferentes bases de dados: o caso da biblioteca do instituto fernandes figueira**. Master's thesis, Fundação Instituto Oswaldo Cruz e Escola Nacional de Saúde Pública, 2005.
- [18] DOS SANTOS, V. **Uma arquitetura suportada por busca semântica para recuperação de fontes de informação em repositórios de metadados**. Master's thesis, Programa de Pós-Graduação em Informática, Universidade Federal do Estado do Rio de Janeiro, 2011.
- [19] DUBLIN.CORE. **Elementos do núcleo de metadata "dublin core", versão 1.1: descrição de referência**. Disponível em: <<http://purl.pt/201/1/>>. Acesso em: 24 fev. 2010, 2010.

- [20] ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. Pearson Addison Wesley, São Paulo, 2005.
- [21] EN.WIKIPEDIA.ORG. **Ralph kimball**. Disponível em <<http://en.wikipedia.org/wiki/RalphKimball>>. Acesso em: 17 mar. 2010, 2010.
- [22] FAPESP. **Protocolo oai-pmh**. Disponível em <<http://clube-oai.incubadora.fapesp.br/portal/tematicas/documentacao>>. Acesso em: 17 dez. 2009, 2009.
- [23] FAPESP. **Protocolo oai-pmh**. Disponível em <<http://clube-oai.incubadora.fapesp.br/portal/prot-oai>>. Acesso em: 17 dez. 2009, 2009.
- [24] FAPESP. **Protocolo oai-pmh**. Disponível em <<http://clube-oai.incubadora.fapesp.br/prot-oai/pratica>>. Acesso em: 17 dez. 2009, 2009.
- [25] FERREIRA, S. M. S. P.; SOUTO, L. F. **Dos sistemas de informação federados à federação de bibliotecas digitais**. *Revista Brasileira de Biblioteconomia e Documentação, Nova Série, São Paulo, 2*, 2006.
- [26] FOUNDATION, A. W. M. **Dublin core metadata initiative**. Disponível em: <<http://dublincore.org/metadata-basics/>>. Acesso em: 23 set. 2010, 2010.
- [27] FOUNDATION, A. W. M. **Open archives initiative**. Disponível em <<http://www.openarchives.org/>>. Acesso em: 21 jul. 2011, 2011.
- [28] GILL, T.; GILLILAND, A. J.; WHALEN, M.; WOODLEY, M. S. **Introduction to metadata**. Disponível em: <<http://www.getty.edu/research/publications/electronicpublications/intrometadata/index.html>>. Acesso em: 05 jan. 2010, 2010.
- [29] GRAVES, M. **Projeto de banco de dados com XML**. Pearson Education do Brasil, São Paulo, 2003.
- [30] HASLHOFER, B.; KLAS, W. **A survey of techniques for archieving metadata interoperability**. *ACM Computing Surveys*, 42, 2010.
- [31] INMON, W. H. **Building the data warehouse**. John Wiley and Son Editors, 1996.
- [32] ISO. **International organization for standardization**. Disponível em: <<http://lcweb.loc.gov/catdir/bibcontrol/caplan.html>>. Acesso em: 23 set. 2011, 2011.
- [33] ISO.IEC.11179. **iso iec 11179. information technology: specification and standardization of data elements**. 1999.

- [34] JAYAKANTH, F.; MALY, K.; ZUBAIR, M.; ASWATH, L. **A dynamic approach to make cds/isis databases interoperability over internet using oai protocol**. Disponível em <www.emeraldinsight.com/journals.htm?articleid=1563492&show=pdf>. Acesso em: 09 set. 2010, 2010.
- [35] JR., T. R.; BANERJEE, K. **Building digital libraries: a how-to-do-it manual**. Neal-Schuman Publishers, Inc, New York/ London, 2008.
- [36] KALOYANOVA, S.; BETTI, G. L.; CASTELLANI, F.; KEIZER, J. **Achieving oai pmh compliancy for cds/isis databases**. *Food and Agriculture Organization of the United Nations, n. 3*, 26, 2008.
- [37] KERHERVE, B.; GERBÉ, O. **Models for metadata or metamodels for data?** *Proc. 2nd IEEE Metadata Conference, Silver*, 1997.
- [38] KOWATA, E. T.; DA SILVA, J. C. **Consulta a banco de dados guiadas por ontologias**. Internal report, Instituto de Informática da Universidade Federal de Goiás, Fevereiro 2011.
- [39] KOWATA, E. T.; DA SILVA, J. C. **Metadados no contexto de banco de dados relacionais: extração e exposição com o protocolo oai-pmh**. Internal report, Instituto de Informática da Universidade Federal de Goiás, Fevereiro 2011.
- [40] LIBRELOTTO, G. R.; RAMALHO, J. C.; HENRIQUES, P. R. **Obtendo interoperabilidade semântica em sistemas heterogêneos de informação com metamorphosis**. Universidade do Minho, Departamento de Informática, Braga, Portugal. Disponível em: <<http://hdl.handle.net/1822/1020>>. Acesso em: 23 set. 2010, 2010.
- [41] MACHADO, M. M. **Open archives: panorama dos repositórios**. Master's thesis, Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, 2006.
- [42] MARCONDES, C. H. S. **Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em c&t: a proposta da biblioteca digital brasileira**. *Ci Inf, Brasília 2001*, 30 (3):24–33, 2001.
- [43] MARCONDES, C. H. S. **Documentos digitais e novas formas de cooperação entre sistemas de informação em c&t**. *Ci Inf, Brasília 2002*, 31 (3):42–54, 2002.
- [44] MARTINS, A. L. **Metadados para descrição de documentos remanescentes de fundo eclesiástico: uma proposta de metadados e biblioteca digital para os assentos de batismo, casamento e óbito da matriz do pilar de ouro preto**. Master's

- thesis, Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Minas Gerais, 2006.
- [45] MASSON, S. M. **Os repositórios digitais no âmbito da sociedade informacional.** *Escola de Comunicação e Artes da Universidade de São Paulo. PRISMA.COM n° 7 - ISSN: 1646-3153*, 2008.
- [46] MILLER, P. **Interoperabilidade: what is it and why should i want it?** Disponível em <<http://www.ariadne.ac.uk/issue24/intro.html>>. Acesso em: 23 nov. 2011, 2011.
- [47] MYSQL. **A diagram of the mysql information schema.** Disponível em <http://www.xcdsql.org/Misc/MySQL_INFORMATION_SCHEMA.html>. Acesso em: 07 jul. 2010, 2010.
- [48] MYSQL. **Information_schema tables.** Disponível em <<http://dev.mysql.com/doc/refman/5.5/en/information-schema.html>>. Acesso em: 07 jul. 2011, 2011.
- [49] MYSQL. **Mysql.** Disponível em <<http://dev.mysql.com/doc/refman/5.0/en/information-schema.html>>. Acesso em: 23 jun. 2011, 2011.
- [50] NISO. **Understanding metadata.** Disponível em <<http://www.niso.org/>>. Acesso em: 21 jan. 2011, 2011.
- [51] OAFORUM. **Oai for beginners: overview.** Disponível em <<http://www.oaforum.org/tutorial>>. Acesso em: 27 maio 2011, 2011.
- [52] OAI-PMH. **Oai-pmh - open archives initiative protocol for metadata harvesting.** Disponível em <<http://www.openarchives.org/OAI/openarchivesprotocol.html/>>. Acesso em: 21 jul. 2011, 2011.
- [53] OAI-PMH. **The open archives initiative protocol for metadata harvesting.** Disponível em <<http://openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: 24 jan. 2011, 2011.
- [54] ORACLE. **Banco de dados oracle.** Disponível em <<http://www.oracle.com>>. Acesso em: 22 dez. 2009, 2009.
- [55] PHPMYADMIN. **Administração do banco de dados mysql.** Disponível em: <<http://phpmyadmin.net/homepage/index.php>>. Acesso em: 24 fev. 2010, 2010.
- [56] POSTGRESQL. **Banco de dados postgresql.** Disponível em: <<http://www.postgresql.org/about/>>. Acesso em: 22 dez. 2010, 2010.

- [57] POSTGRESQL. **Postgresql: catálogo do sistema.** Disponível em: <<http://pgdocptbr.sourceforge.net/pg80/catalogs.html>>. Acesso em: 24 dez. 2010, 2010.
- [58] PRESSMAN, R. S. **Engenharia de Software.** McGraw-Hill, São Paulo, 2006.
- [59] PT.WIKIPEDIA.ORG. **Interoperabilidade.** Disponível em <<http://pt.wikipedia.org/wiki/interoperabilidade>>. Acesso em: 21 set. 2010, 2010.
- [60] RAN PARK, J.; TOSAKA, Y. **Metadata creation practices in digital repositories and collections: schemata, selection criteria, and interoperability.** *Information technology and libraries, September, 2010.*
- [61] REPOSITORY.EXPLORER. **Repository explorer.** Disponível em <<http://re.cs.uct.ac.za>>. Acesso em: 05 jun. 2011, 2011.
- [62] SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistemas de banco de dados.** Elsevier, Rio de Janeiro, 2006.
- [63] SÁNCHEZ, J. A.; RAZO, A.; CÓRDOVA, J. M.; VILLEGAS, A. **Dynamic generation of oai servers.** *Joint Conference on Digital Libraries (JCDL)*, p. 258–259, 2006.
- [64] TAMMARO, A. M.; SALARELLI, A. **A biblioteca digital.** Brinquet de Lemos Livros, Brasília, 2008.
- [65] UNIVERSITY.OF.QUEENSLAND.LIBRARY. **Metadata: an introduction to metadata.** Disponível em: <<http://www.library.uq.edu.au/iad/ctmeta4.html>>. Acesso em: 21 jan. 2010, 2010.
- [66] W3C. **Extensible markup language (xml).** Disponível em <<http://www.w3.org/XML>>. Acesso em: 28 jul. 2011, 2011.
- [67] WITTEN, I. H.; BAINBRIDGE, D.; NICHOLS, D. M. **How to build a digital library.** Morgan Kaufmann, United States, 2003.

Guia de Instalação

Este apêndice apresenta um "guia de instalação do aplicativo", que traz a relação de tecnologias utilizadas, bem como a orientação para configurar o ambiente, o banco de dados e o servidor *web* para a execução do aplicativo. Dessa forma, as tecnologias utilizadas estão relacionadas na Seção [A.1](#), a configuração do ambiente na Seção [A.2](#), a configuração do banco de dados na Seção [A.3](#), a instalação e a configuração do aplicativo na Seção [A.4](#).

A.1 Tecnologias Empregadas

Para o desenvolvimento do *METADATABASE EXTRACTOR* foram utilizados os seguintes *softwares*: PHP 5.2.6, Apache 2.0.63, Mysql 5.0.51b, PhpMyAdmin 2.11.7 e para a navegação e publicação: *Mozilla Firefox* e *Filezilla Client*, respectivamente.

Instalar no servidor *web* o PHP 5.2.6 ou posterior, Apache 2.0.63, MySQL 5.0.51b e o PhpMyAdmin 2.11.7. O PhpMyAdmin é utilizado para gerenciar o banco de dados, o *Mozilla Firefox* é o navegador padrão do aplicativo e o *Filezilla Client* é utilizado para transferência/publicação de arquivos na *web*.

A.2 Configuração do Ambiente

Para instalar o aplicativo é necessário ter um domínio ou subdomínio que será o endereço do aplicativo, por exemplo: www.metadatabaseextractor.com.br ou www.unucseh.ueg.br/metadatabase.

A.3 Configuração do Banco de Dados

Nesta seção encontram-se as configurações básicas de banco de dados para a execução do aplicativo. Os passos são:

1. Criar o banco de dados: METADATABASE_EXTRACTOR (Se desejar, utilize um banco de dados existente. Na configuração do sistema altere o banco de dados metadatabase por um que queira utilizar);
2. Criar as tabelas: metadbx_tb_oai_records - Código A.1, metadbx_cad_admdatabase - Código A.2 e metadbx_relatorios_bancodados - Código A.3. A estrutura das tabelas estão definidas na Seção A.5.

Os arquivos com as estruturas das tabelas em *Structured Query Language* (SQL) encontram-se no CD de instalação e também no site www.unucseh.ueg.br/metadatabase. Os detalhes das tabelas são descritas a seguir.

Os conteúdos dos bancos de dados e das bibliotecas digitais foram registrados na tabela metadbx_tb_oai_records, e a estrutura está apresentada no Código A.1.

Os dados dos administradores do sistema são armazenados na estrutura apresentada no Código A.2 e o dados de relatórios de bancos de dados disponíveis são armazenados na estrutura apresentada no Código A.3.

A.4 Instalação e Configuração do Aplicativo

Com o *Filezilla Client*, transfira os arquivos que se encontram no CD (pasta sistema) para o domínio/subdomínio criado na Seção A.2.

Localize os arquivos `conexao.php`, `conexao_sys.php` na pasta `sistema/source/conecta`. Configure o acesso ao banco de dados, informando o nome do servidor, usuário e a senha do banco de dados.

Um exemplo de configuração da conexão com o banco de dados MySQL do servidor é: `$conexao=mysql_connect("servidor","usuario","senha")`.

A.5 Definição das Estruturas das Tabelas em SQL

Código A.1 – Código SQL - Código SQL da tabela metadbx_tb_oai_records

```
1  --
2  -- Estrutura da tabela 'metadbx_tb_oai_records'
3  --
4
5  CREATE TABLE IF NOT EXISTS `metadbx_tb_oai_records` (
6    `serial` int(11) NOT NULL AUTO_INCREMENT,
7    `provider` varchar(255) DEFAULT NULL,
8    `url` varchar(255) DEFAULT NULL,
9    `email` varchar(255) DEFAULT NULL,
10   `oai_set` varchar(255) DEFAULT NULL,
11   `dispquery` enum('false','true') NOT NULL
12   COMMENT 'Disponível para consulta externa',
13   `dc_title` varchar(255) DEFAULT NULL,
14   `dc_creator` text,
15   `dc_subject` varchar(255) DEFAULT NULL,
16   `dc_description` text,
17   `dc_contributor` varchar(255) DEFAULT NULL,
18   `dc_publisher` varchar(255) DEFAULT NULL,
19   `dc_date` date DEFAULT NULL,
20   `dc_type` varchar(255) DEFAULT NULL,
21   `dc_format` varchar(255) DEFAULT NULL,
22   `dc_identifier` varchar(255) DEFAULT NULL,
23   `dc_source` varchar(255) DEFAULT NULL,
24   `dc_language` varchar(255) DEFAULT NULL,
25   `dc_relation` varchar(255) DEFAULT NULL,
26   `dc_coverage` varchar(255) DEFAULT NULL,
27   `dc_rights` varchar(255) DEFAULT NULL,
28   `loginbd` varchar(15) NOT NULL
29   COMMENT 'Login do banco de dados',
30   `passwordbd` varchar(15) NOT NULL
31   COMMENT 'Senha de acesso ao banco de dados',
32   PRIMARY KEY (`serial`)
33 ) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=1;
```

Código A.2 – Código SQL - Código SQL da tabela metadbx_cad_admdatabase

```

1  --
2  -- Estrutura da tabela `metadbx_cad_admdatabase`
3  --
4
5  CREATE TABLE IF NOT EXISTS `metadbx_cad_admdatabase` (
6    `cod_adm` int(10) NOT NULL AUTO_INCREMENT,
7    `nome_adm` varchar(60) COLLATE utf8_unicode_ci NOT NULL
8    COMMENT 'nome do administrador de BD',
9    `login_adm` varchar(15) COLLATE utf8_unicode_ci NOT NULL,
10   `senha_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
11   `cpf_adm` varchar(14) COLLATE utf8_unicode_ci NOT NULL
12   COMMENT 'cpf do administrador de BD',
13   `data_nascimento_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
14   `sexo_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL,
15   `email_adm` varchar(60) COLLATE utf8_unicode_ci NOT NULL,
16   `data_cadastro_adm` varchar(10) COLLATE utf8_unicode_ci NOT NULL
17   COMMENT 'Data de cadastro',
18   `acesso_adm` int(1) NOT NULL
19   COMMENT 'Tipo de acesso ao sistema (9-administrador do sistema ou
20    1-administrador do banco de dados ou 0-usuario)',
21   PRIMARY KEY (`cod_adm`)
22 ) ENGINE=MyISAM DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci
23 COMMENT='Cadastro de Administradores de Banco de Dados' AUTO_INCREMENT=1;

```

Código A.3 – Código SQL - Código SQL da tabela metadbx_cad_relatorio_bancodados

```

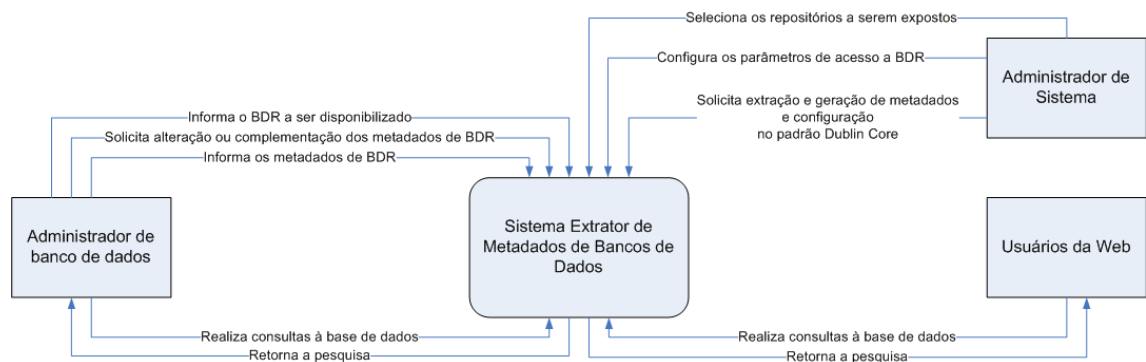
1  --
2  -- Estrutura da tabela `metadbx_relatorio_bancodados`
3  --
4
5  CREATE TABLE IF NOT EXISTS `relatorios_bancodados` (
6    `cod_rel` int(10) NOT NULL auto_increment,
7    `query_rel` varchar(60) collate utf8_unicode_ci NOT NULL
8    COMMENT 'Descrição da consulta a banco de dados',
9    `sql_rel` varchar(250) collate utf8_unicode_ci NOT NULL
10   COMMENT 'linha de consulta sql a ser submetido ao banco de dados',
11   `serial_bd` int(10) NOT NULL,
12   PRIMARY KEY (`cod_rel`)
13 ) ENGINE=MyISAM DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci
14 COMMENT='Cadastro Relatórios' AUTO_INCREMENT=1 ;

```

Diagrama de Contexto e Diagrama de Entidade Relacionamento

Este apêndice apresenta o diagrama de contexto e o diagrama de entidade relacionamento do sistema extrator de metadados de bancos de dados relacionais.

B.1 Diagrama de Contexto do Sistema



B.2 Diagrama de Entidade Relacionamento

