AT-2398

Uma arquitetura de informática para integração de sistemas de bibliotecas na Internet*

MSc. Geórgia R. R. Gomes – Dr. Rubens N. Melo – MSc. Sérgio da C. Côrtes

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

{georgia@dbd.puc-rio.br,rubens@inf.puc-rio.br,scortes@inf.puc-rio.br}

Resumo

Este trabalho apresenta uma arquitetura de informática utilizando a tecnologia de Banco de Dados e de Engenharia de Software que tem como objetivo integrar diferentes sistemas de bibliotecas eletrônicas e digitais na Internet criando uma biblioteca virtual.

Palavras chaves: Biblioteca Virtual, Internet, Banco de Dados Heterogêneos, USMARC, Dublin Core.

1. Introdução

Sistemas de Bibliotecas são um importante instrumento em todos os níveis de pesquisa, seja ela científica ou não. Integrar esses sistemas, proporcionando a seus usuários facilidades na busca de referências bibliográficas e, em muitos casos, da própria obra, é uma necessidade cada vez maior. Integrar, disponibilizar e visualizar dados é uma área de pesquisa bastante intensa em informática, principalmente em Banco de Dados.

* Este trabalho é baseado na dissertação de mestrado." *Um ambiente para integração de Dados Bibliográficos baseado em Mediadores*", Departamento de Informática, PUC-Rio.

INFOBILA

Este trabalho apresenta e desenvolve uma arquitetura de informática, utilizando a tecnologia de Banco de Dados e de Engenharia de Software, bem como um protótipo de funcionamento na WEB, capaz de integrar diferentes bibliotecas digitais, disponibilizando para seus usuários uma biblioteca virtual, onde o usuário, através de um equipamento plugado a Internet ou Intranet, pesquisa e consulta documentos e referências bibliográficas em diferentes instituições.

O pioneirismo desse trabalho se dá na utilização das tecnologias de informática, principalmente de banco de dados, engenharia de software, comunicação de dados e visualização de dados, buscando uma solução integradora para a área de sistemas de bibliotecas e implementando um protótipo para provar a eficácia da arquitetura proposta.

Entre os objetivos principais podemos destacar: promover a integração entre sistemas de bibliotecas desenvolvidos em ambientes e plataformas de informática completamente diferentes e heterogêneos; possibilitar a utilização de acervos bibliográficos de instituições hoje isoladas, integrando-os ao ambiente da Internet; disponibilizar para o público em geral um acervo bibliográfico ilimitado; promover o intercâmbio do conhecimento científico entre as áreas de informática e de bibliotecas; disponibilizar tecnologias de ponta da informática para sustentar a área de biblioteca; apresentar uma solução de custos reduzidos para integração de bibliotecas.

O trabalho está organizado da seguinte forma: a seção 2,apresenta as definições conceituais de bibliotecas, a seção 3 descreve metadado e alguns padrões utilizados para catalogação de dados bibliotegráficos de metadados (USMARC e Dublin Core), a seção 4 apresenta a arquitetura proposta para integração de bibliotecas e a seção 5 apresenta a conclusão do trabalho.

2. Definição de Biblioteca

Segundo [Barker94] existem quatro definições de bibliotecas: Convencional (polimídia), eletrônica, digital e virtual. Nos itens a seguir, descreveremos, de forma resumida essas definições, somente buscando fornecer subsídios para os leitores desse trabalho.

2.1 – Biblioteca Convencional

Atualmente diversas mídias são utilizadas como meios independentes para armazenamento da informação. As bibliotecas convencionais possuem livros e periódicos que convivem com fitas, vídeos, CD-ROMs, microfilmes, softwares de controle e armazenamento etc, recebendo, também, desta forma o nome de bibliotecas polimídias. Os processos de gerenciamento e organização nestas bibliotecas são praticamente manuais e, apesar dos computadores estarem disponíveis para os usuários, esta tecnologia não é utilizada para a realização de qualquer forma de automação das bibliotecas.

2.2 – Biblioteca Eletrônica

A biblioteca Eletrônica se refere a sistemas de bibliotecas, nos quais os processos básicos da biblioteca são de natureza eletrônica, o que implica ampla utilização de computadores e de suas facilidades na construção de índices on-line, busca de textos completos e na recuperação e armazenagem de registros.

2.3 – Biblioteca Digital

A biblioteca Digital contém apenas informação na forma digital, podendo residir em meios diferentes de armazenamento, como memórias eletrônicas, tais como os discos magnéticos e óticos. Desta forma, a biblioteca digital não contém livros na forma

convencional. A informação pode ser acessada em locais específicos e/ou remotamente, por meio de redes de computadores.

2.4 – Biblioteca Virtual

A biblioteca Virtual é conceituada como um tipo de biblioteca que, para existir, depende da tecnologia da realidade virtual. Neste caso, um software próprio, acoplado a um computador sofisticado, reproduz o ambiente de uma biblioteca em duas ou três dimensões, criando um ambiente de total imersão e interação. É então possível, ao entrar em uma biblioteca virtual, circular entre as salas, selecionar um livro nas estantes, "tocá-lo", "abri-lo" e "lê-lo". Obviamente, o único "lugar" onde o livro realmente existe é no computador e na imaginação do leitor. Para pesquisadores da área de Ciência da Informação este conceito não é único, podendo também possuir as seguintes definições:

- Acesso por meio de redes a recursos de informação disponíveis em sistemas de base computadorizadas, normalmente remotos [Poulter94];
- Conjunto com utilização de recursos eletrônicos. Disto deriva o fato de a expressão "biblioteca virtual" ser utilizada como um sinônimo de "biblioteca eletrônica", ou ainda *desktop library* [Kemp94], [Deschamps94], [Cloyes94].
- Acesso remoto aos conteúdos e serviços de bibliotecas e outros recursos de informação, combinando uma coleção interna de materiais correntes e fartamente usados em ambas as formas (eletrônica e impressa), com redes eletrônicas que provêm acesso e a transferência de fontes de conhecimento e de informação, com bibliotecas e instituições comerciais externas em todo o mundo [Piggott93] , [Saunders95].

Com estas definições, conceituamos Biblioteca Virtual como um conjunto de documentos e dados bibliográficos armazenados em bases digitais, com acesso por meio de redes, e que não existe fisicamente.

Os sistemas de bibliotecas têm crescido muito e todas as bibliotecas polimídias estão se adequando às novas tecnologias. O objetivo dos usuários de uma biblioteca é encontrar uma obra, estando ela onde estiver, e o objetivo dos mantenedores de bibliotecas é disponibilizar suas obras para um número cada vez maior de pessoas.

3. Padrões de Metadados

Segundo [Sumpter94], "Metadado é a informação sobre o dado que permite o acesso e gerenciamento deste dado de maneira eficiente e inteligente."

A palavra metadados foi criada por Jack Myres, em 1969, para denominar os dados que descreviam registros de arquivos convencionais.

Metadados se aplicam a uma grande variedade de acervos de dados convencionais que podem, ou não, estar disponíveis em redes eletrônicas de computadores, tais como acervos de dados bancários, de bibliotecas tradicionais ou acervos de dados não convencionais como os de sistemas de informações geográficas, de bibliotecas digitais, documentos multimídia etc.

A especificação e utilização de padrões garantem a existência de um conjunto de informações comuns sobre um determinado tema ou área, com regras claramente estabelecidas e aceitas pela comunidade envolvida. Padrões facilitam a compreensão, a integração e o uso compartilhado de informações entre usuários de diferentes formações, com diferentes níveis de experiências e diferentes propósitos. O estabelecimento de padrões implica o compromisso entre usuários e provedores de informações, que devem mutuamente aceitar, colaborar e usar as terminologias e definições estabelecidas .

Um padrão de metadados é formado por um conjunto de elementos descritores que podem estar relacionados. Geralmente são padronizados nomes, informações ou grupos de dados utilizados para descrever um determinado tipo de acervo. Caso existam relacionamentos, estes também devem constar do modelo de padronização. Geralmente, a definição de padrões de metadados é feita por um grupo de pessoas onde, entre seus componentes, se encontram usuários que detêm o conhecimento sobre um determinado tipo de acervo. No nosso caso, bibliotecários ou profissionais de ciência da informação junto com profissionais de informática.

O grau de complexidade de um padrão de metadados pode ser elevado quando, no ambiente a ser descrito, existe uma grande diversidade de informações manipuladas, cada qual com características diferentes e que devem ser integradas de forma a se obter um modelo de padronização coerente.

Apesar da complexidade de alguns padrões de metadados, o conjunto de descritores deve conter apenas informações apropriadas e suficientes para descrever o dado de forma que a informação nele contida além de ser compreendida por qualquer pessoa, possa também ser compilada/interpretada pelo computador, pois pode servir de subsídio a sistemas de busca e recuperação de informações[Yeager96].

A seguir apresentaremos dois padrões de metadados USMARC e Dublin Core que são utilizados para descrever dados bibliográficos.

3.1 – USMARC

Nos anos 60 a Biblioteca do Congresso Americano (Library of Congress - LC) criou o formato LC MARC , um sistema de uso conciso de números, letras e símbolos dentro do próprio registro bibliográfico, com a finalidade de catalogar diferentes tipos de informações.

O LC MARC original evoluiu e se tornou o padrão USMARC, que é utilizado pela maioria dos sistemas de bibliotecas.

O USMARC - **MA**chine-**R**eadable **C**ataloging significa registro de catalogação legível por máquina.

MAchine-Readable – Legível por máquina, um computador pode ler e interpretar os dados do registro de catalogação.

Cataloging – Registro de Catalogação, significa um registro bibliográfico, ou a informação tradicionalmente mostrada em uma ficha de um catálogo de biblioteca. O registro inclui (não necessariamente nesta ordem): descrição do item, entrada principal e entradas adicionais (secundárias), cabeçalho de assunto, classificação ou número de chamada. Um registro MARC contém, freqüentemente, muitas informações adicionais.

- Descrição do item – Os Bibliotecários geralmente utilizam a *Anglo-American Cataloguing Rules* (regras de Catalogação Anglo Americana), 2nd ed., 1988 revisada , popularmente conhecida como AACR2R, para compor a descrição bibliográfica de um item de biblioteca. Inclui o título, declaração de responsabilidade, edição, detalhes específicos de material, informação de publicação, descrição física, séries, notas e números padrões.
- Entrada principal e entradas adicionais (secundária) – AACR2R também contém regras para determinar "pontos de acesso" para o registro, normalmente chamada "entrada principal" e "outras entradas adicionais" (entradas secundárias) , e a forma que estes pontos de acessos devem tomar. Pontos de acesso são pontos de recuperação no catálogo da biblioteca onde os usuários seriam capazes de ver o item.

  As regras da AACR2R são usadas para responder questões como: Para este livro, deveria haver entrada no catálogo para mais de um autor ou mais de um título?

Como o nome do autor deveria ser escrito? Este título é uma entrada principal?(no caso de não haver um autor).

- Cabeçalho de Assunto – Os Bibliotecários utilizam normalmente, listas de cabeçalhos de assunto, podendo ser a da LC ou alguma outra lista, para selecionar o assunto relacionado ao item. O uso de uma lista é importante para a consistência e assegurar que serão encontrados todos os itens referentes a um determinado assunto. Por exemplo, a lista de cabeçalho de assunto indica que todos os livros sobre gatos devem ser associados ao assunto GATOS. Usando este cabeçalho autorizado eliminamos a possibilidade de listar alguns livros sob o assunto GATOS e outros sob o assunto FELINOS. Até mesmo se um livro é chamado *Tudo sobre Felinos*, o cabeçalho de assunto será GATOS. Deste modo todos os livros deste assunto ficarão reunidos sob um mesmo assunto/termo/palavra, não sendo necessário imaginar os possíveis sinônimos da palavra para se encontrar o que deseja.
- Número de Chamada – Os Bibliotecários utilizam a classificação decimal de Dewey ou a classificação decimal universal ou outra classificação específica para selecionar o número de chamada do item. A finalidade do número de chamada é colocar itens dos mesmos assuntos juntos na estante da biblioteca. A Segunda parte de um número de chamada normalmente representa o nome do autor, facilitando a sub-classificação (cutter).

Com a utilização do padrão USMARC as bibliotecas evitam duplicação de trabalho, podendo compartilhar os registros bibliográficos. A maioria dos sistemas de bibliotecas utilizam o padrão USMARC, facilitando desta forma a integração, intercâmbio, o acesso e a interpretação de dados.

O formato USMARC portanto é um conjunto de códigos e designações de conteúdos definido para codificar registros legíveis por máquina. São definidos formatos para cinco tipos de dados: bibliográficos, propriedades, autoridades, classificação e informação de comunidade. Nesta dissertação estudaremos o formato dos dados bibliográficos.

O formato USMARC para dados bibliográficos contém especificações para codificar elementos que precisam ser descritos, recuperados e controlados; é um formato integrado e definido para identificar e descrever formas diferentes de material bibliográfico.

3.1.1 – Organização do registro USMARC

Um registro USMARC apresenta três seções principais: o líder (*leader)*, o diretório (*directory*) e os campos variáveis (*variable fields*).

3.1.1.1 – Líder

Consiste em elementos de dados que contêm valores codificados e são identificados através da posição do caracter relativo. Elementos de dados no líder definem parâmetros para processar o registro. O líder tem um tamanho fixo de 24 caracteres e ocorre no inicio de cada registro.

3.1.1.2 – Diretório

Contém tags (etiquetas), que indicam o inicio da localização de cada campo e o seu tamanho no registro. Primeiro aparecem as variáveis dos campos de controle, depois em ordem as tags. Seguem as entradas para campos de dados variáveis, organizadas em ordem ascendente, de acordo com o primeiro caracter da tag. A ordem dos campos no registro não corresponde, necessariamente, à ordem de entrada de diretório. Tags duplicatas são identificadas somente pela localização dos respectivos campos dentro do registro. O tamanho

da entrada do diretório é definido nas posições 20 a 23 do líder. No formato USMARC cada entrada do diretório tem o tamanho de 12 caracteres. O diretório termina com o caracter de fim de campo.

3.1.1.3 – Campos Variáveis

O conteúdo de dados de um registro é dividido em campos variáveis. Os formatos de USMARC distinguem dois tipos de campos variáveis: campos de controle variáveis e campos de dados variáveis. Campos de controle e campos de dados são distintos somente por estrutura. O termo campo fixo é ocasionalmente usado na documentação do USMARC , referindo-se também a campos de controle ou a campos específicos de dados de controle, por exemplo, 007 (descrição física do campo fixo) ou 008 (fixed-length data elements).

3.1.2 – Regras Gerais.

Cada registro bibliográfico é dividido logicamente em campos. Há um campo para autor, um para título e assim por diante. Cada campo é identificado por uma tag de três caracteres. Estes campos são subdivididos em um ou mais subcampos.

Campos variáveis são agrupados em blocos, de acordo com o primeiro caracter da tag, que identifica a função dos dados dentro do registro, por exemplo, entrada principal, entrada secundária, entrada de assunto. O tipo de informação no campo, por exemplo, nome pessoal, nome de corporações ou título é identificado nos caracteres restantes da tag.

Há algumas regras gerais que ajudam a definir o significado dos campos de cada tag. Observe que na discussão das tags USMARC, a notação XX é freqüentemente usada para referenciar um grupo de tags. Por exemplo, 1XX se refere a todas as tags que estão no grupo 100: 100, 110, 130 e assim por diante.

3.1.2.1 – Blocos do formato bibliográfico.

As tags são divididas por cem. A divisão básica de um registro bibliográfico é:

0XX – Informações de controle, números e códigos

1XX – Entrada principal

2XX – Títulos, edição, impressão

3XX – Descrição física, etc.

4XX – Declaração de séries

5XX – Notas

6XX – Assunto

7XX – Entradas secundárias

8XX – Entradas secundárias de séries, localizações, etc

Os 9XXs são usados para definições de uso local, como numero de código de barras. Os grupos X9X, 09X, 59X, etc. também são usados para uso local, excluindo o 490 (declaração de séries).

3.1.2.2 – Conteúdos Paralelos.

No USMARC temos a designação dos conteúdos paralelos, isto é, os dois digitos finais da tag são complementados com os blocos abaixo:

X00 – Nomes pessoais

X10 – Nomes de corporações

X11 – Nomes de Congressos

X30 – Título uniforme

X40 – Títulos bibliográficos

X50 – Termos tópicos

X51 – Nomes geográficos.

Por exemplo, temos a entrada principal (1XX) que será um nome pessoal (X00), então a tag será 100. Para assunto (6XX) nome pessoal (X00) a tag é 600 e assim por diante.

3.1.2.3 – Indicadores.

Depois de cada tag temos duas posições (com exceção dos campos 001 à 009), onde são definidos os indicadores. Em alguns campos utiliza-se a primeira ou a segunda posição, em alguns as duas posições e em outros como o 020 e 300, nenhuma posição é utilizada. Quando uma posição de indicador não é usada este é chamado de indefinido, utilizando-se o caracter # para representá-lo.

Os indicadores podem ser caracteres alfabéticos ou numéricos. Primeiro são definidos valores numéricos. Cada valor de indicador é um numero de 0 à 9. O valor 9 é reservado para implementação local. No exemplo abaixo os três primeiros dígitos são a tag (245 define como um campo de título) e os próximos dois dígitos (1 e 4) são valores de indicador. O 1 é o primeiro indicador e o 4 é o segundo indicador.

245 1 4 $a The emperor's new clothes / $c adapted from
Hans Christian Andersen and illustrated by Janet Stevens.

O primeiro indicador contém um valor que especifica se um título tem entrada secundária (adicional) ou não.

O valor 1 no primeiro indicador no campo título indica que deveria ser separada a entrada de título no catalogo. No ambiente de catálogo de fichas isto significa que uma ficha de título deveria ser impressa para este item e uma entrada de título secundário deveria ser mostrada. O valor 0 no primeiro indicador significaria que a entrada principal é o próprio título.

3.1.2.4 – Subcampos.

A maioria dos campos contém várias partes de dados. Cada tipo de dado dentro do campo é chamado de subcampo e cada subcampo é precedido por um código de subcampo, estes identificam elementos de dados dentro de um campo que requer (ou poderia requerer) manipulação separada. Campos 001 a 009 não contém subcampos.

Códigos de subcampos são representados por uma letra minúscula (ocasionalmente um número) que precede um delimitador. Um delimitador é um caracter usado para separar subcampos. Cada código de subcampo indica qual o tipo de dado que o segue. Para cada campo na documentação do formato USMARC bibliográfico é definida uma lista de códigos de subcampos válidos.

No exemplo abaixo, o campo para descrição física de um livro (definida na tag 300) inclui um subcampo para a extensão (número de páginas), um subcampo para outros detalhes físicos (informação de ilustração) e um subcampo para dimensões (centímetros):

300 ## $a 675 pág.: $b ill.; $c 24 cm.

No exemplo acima, os códigos de subcampos são: $a para a extensão, $b para outros detalhes físicos e $c para dimensão.

Os softwares de bibliotecas usam diferentes caracteres para representar os delimitadores. Por exemplo (@), ($), (_). Nesta dissertação usamos o símbolo $.

Este padrão é o mais utilizado para catalogação de dados bibliográficos.

3.2 – Dublin Core

Com o objetivo de reunir esforços para alcançar uma solução comum para o problema de localização de informação na Internet, organizações como a OCLC (Online Computer Library Center) e o NCSA (National Center for Supercomputing Aplications)

INFOBILA

organizaram, em março de 1995, um workshop sobre metadados em Dublin, Irlanda. Participaram deste workshop profissionais da área de Biblioteconomia, Ciências da Informação, Informática e Provedores de Informações da rede. O resultado foi a definição do padrão de metadados para Internet chamado Dublin Core [Weibel95].

Devido ao tamanho, complexidade e variedade das informações disponíveis na rede, foi definido um limite no escopo do que seria descrito. Como a maioria dos objetos na Internet está em forma de documentos, o objetivo principal do Dublin Core foi identificar e definir um conjunto contendo o mínimo de elementos capazes de descrever "Objetos do Tipo Documento" (ou Documents Like Objects – DLOs) da Internet. Este conjunto teve que ser suficientemente simples para ser entendido e usado por um grande números de autores e provedores que contribuem com informações na Internet. Um documento na Internet pode ser composto por textos com chamadas para imagens, vídeos, audio ou outros documentos hipertextos. Não foi imposta nenhuma restrição quanto à composição de um DLO, mas definiu-se que o primeiro recurso a ser tratado seria o texto.

Inicialmente foi definido, pelos participantes do workshop, um conjunto de treze elementos de metadados, chamado *Dublin Core Metadata Set (*ou *Dublin Core),* considerados fundamentais para a descrição dos DLOs, podendo cada um deles conter os qualificadores esquema e tipo que serão descritos a seguir:

- *Type* (tipo) – é usado para generalizar ou especificar a definição semântica de um elemento do Dublin Core. Através do qualificador *Type*, por exemplo, o elemento "autor" pode ser complementado com outras informações relacionadas ao autor, tais como: seu número de telefone, email, fax, etx. Por exemplo: Criador (type=email) = georgia@dbd.puc-rio.br;
- *Schema* (esquema) – representa uma forma de introduzir padronização ao conteúdo dos elementos do padrão Dublin Core, possibilitando que o valor de um elemento seja

interpretado segundo o esquema (de codificação, de classificação, etc.) por ele especificado. Por exemplo: Título (esquema=AACR2) = "Banco deDados em Aplicações Cliente-Servidor".

3.2.1 – Elementos do Dublin Core

Com a finalidade de descrever melhor os objetos e suportar definições que são bastante precisas para mapeamento de registros largamente usados como padrões, tais como USMARC, TEI ou FGDC, cada elemento do Dublin Core pode ser qualificado com um esquema. Esquemas são usados sempre que é necessário descrever a razão para a codificação dos dados associados ao elemento.

- **Assunto**: é o área de conhecimento ao qual o trabalho pertence.
- **Título:** é definido como o nome do objeto.
- **Autor:**é o principal responsável pelo conteúdo intelectual do trabalho.
- **Editor:**é definido como o agente ou agencia responsável pela disponibilização do objeto.
- **Outro Agente**: são pessoas ou organizações diferentes de autores e editores, que deram uma significante contribuição intelectual ao trabalho, tais como tradutores, ilustradores.
- **Data:** reflete a data em que o objeto foi disponibilizado em sua forma atual.
- **Tipo do Objeto:** é definido como uma categoria abstrata ou gênero do objeto, como romance, poema, dicionário, enciclopédia, software executável, código fonte, arquivos de dados ou qualquer outra categoria que se julgue útil para recuperação.
- **Forma**: é definida como representação do dado do objeto, tais como arquivo *postscript*, arquivo executável do windows, arquivo HTML.

- **Identificador:** é a *string* ou número usado para identificar univocamente um objeto.
- **Relação:** identifica o relacionamento do objeto com outros objetos, impressos ou eletrônicos, como outras partes de um documento hierarquizado, outras partes de uma coleção de documentos ou uma série de documentos.
- **Fonte:** objetos, em papel ou meio magnético, do qual este objeto é derivado.
- **Idioma:** especifica o idioma do conteúdo intelectual do objeto descrito
- **Extensão: c**aracterísticas de localização espacial e duração temporal do objeto

A seguir é mostrado um exemplo de descrição de um documento utilizando elementos do padrão Dublin Core:

- Título: Banco deDados em Aplicações Cliente-Servidor
- Autor: Rubens Nascimento Melo
- Autor: Sidney Dias da Silva
- Autor: Asterio K. Tanaka
- Editor: Infobook
- Outro Agente: Tereza Cristina B. de Castro Barbosa
- Data: 1997.
- Objeto: Livro.
- Assunto: Banco de Dados
- Idioma: Português.
- Identificador: 86234

Este padrão, muito discutido nas publicações sobre o assunto, é considerado um marco nas discussões sobre padrões de metadados na Internet e ponto de partida para os demais projetos envolvendo especificação de uma arquitetura de metadados para a Web . Apesar de ser extremamente simples, seu projeto é fruto da grande experiência de

profissionais envolvidos com a catalogação de recursos. É certo que o padrão MARC teve grande influência em seu desenvolvimento, porém foram extraídos desse padrão somente aqueles elementos essenciais à descrição e ao acesso de uma grande variedade de recursos de informação. Sua simplicidade é fator chave para a rápida utilização na Web, embora já existam propostas para a utilização do padrão fora desse contexto na forma de uma aplicação integrando todos os tipos de recursos informação, inclusive aqueles não disponíveis no meio eletrônico

Na tabela a seguir faremos algumas comparações dos Metadados apresentados.

Aspectos Considerados	**MARC**	**Dublin Core**
Comunidade Utilizadora do Padrão	- Bibliotecários - Arquivistas - Museus, etc.	- Bibliotecários - Autores de documentos da WEB em geral
Facilidade de Utilização do Padrão	- Exige treinamento prévio e conhecimento especializado	- Não exige treinamento prévio, nem conhecimento especializado
Cobertura do Padrão	- Documentos, vídeos, gravações sonoras, mapas, periódicos, etc.	- Objetos do tipo documentos (DLOs)
Sintaxe / formato para transporte dos Metadados	- ISO 2709	- Não tem sintaxe definida
Estrutura dos Elementos	- Resgistros (formato interno binário)	- Pares (nome-atributo, valor-atributo)
Utilização de regras de conteúdo para os Elementos de dados	- AACR2, ISBD, Dewey,etc. (uso obrigatório)	- Podem ser usadas (não são obrigatórias)
Protocolos Associados	Z39.50	Z39.50

Tabela 1 - Tabela de comparação dos padrões de Metadados

4. Arquitetura proposta para integração de bibliotecas

A arquitetura que propomos compõe-se de 4 camadas, sendo uma camada de Aplicações, uma camada de Mediadores, uma camada de Tradutores e uma camada de Fontes de Dados Bibliográficos (**figura 1**).

A camada de Aplicações pode ser composta por diferentes tipos de aplicações, específicas de cada ambiente operacional, possuindo as mesmas interfaces de caracteres ou gráficas no ambiente WEB e cliente/servidor (em duas camadas), podem ser desenvolvidas com softwares do tipo C^{++}, Delphi, Java, Visual Basic, etc. Nesta camada o usuário coloca a sua pergunta, faz a sua consulta. No nosso trabalho temos dois tipos de aplicação uma cliente/servidor que chamamos de Cliente Local e um programa para Web onde a camada de aplicação se divide em browser e servidor Web.

A camada de Mediadores, através de sua linguagem padrão para aplicações, interpreta e traduz as consultas dos usuários, para acessar fontes de dados de forma transparente e uniforme. O Mediador transforma as consultas em sub-consultas e as envia, através de uma codificação padrão, para o tradutor correspondente de cada Fonte de Dados. No nosso trabalho criamos 3 mediadores o mediador Autor que faz a consulta por autor o mediador Assunto que faz a consulta por assunto e o mediador Título que faz a consulta por título.

A camada de Tradutores recebe as sub-consultas do mediador e as transforma em consultas na linguagem correspondente de cada Fonte de Dados. O resultado da consulta é enviado ao mediador para que este possa centralizar as respostas dos diversos Tradutores, executar as operações necessárias (União, Interseção, Diferença, Seleção ou Projeção), então,, após enviá-las aos usuários que realizaram as consultas. Neste trabalho criamos dois Tradutores um tradutor para a Biblioteca PUC-Rio onde utilizamos o padrão USMARC e um

tradutor para a biblioteca do Futuro (biblioteca fictícia) onde utilizamos o padrão Dublin Core.

A camada Fontes representam as Fontes de Dados Bibliográficas que contem todos os modelos de dados e os respectivos dados, correspondentes a cada biblioteca.

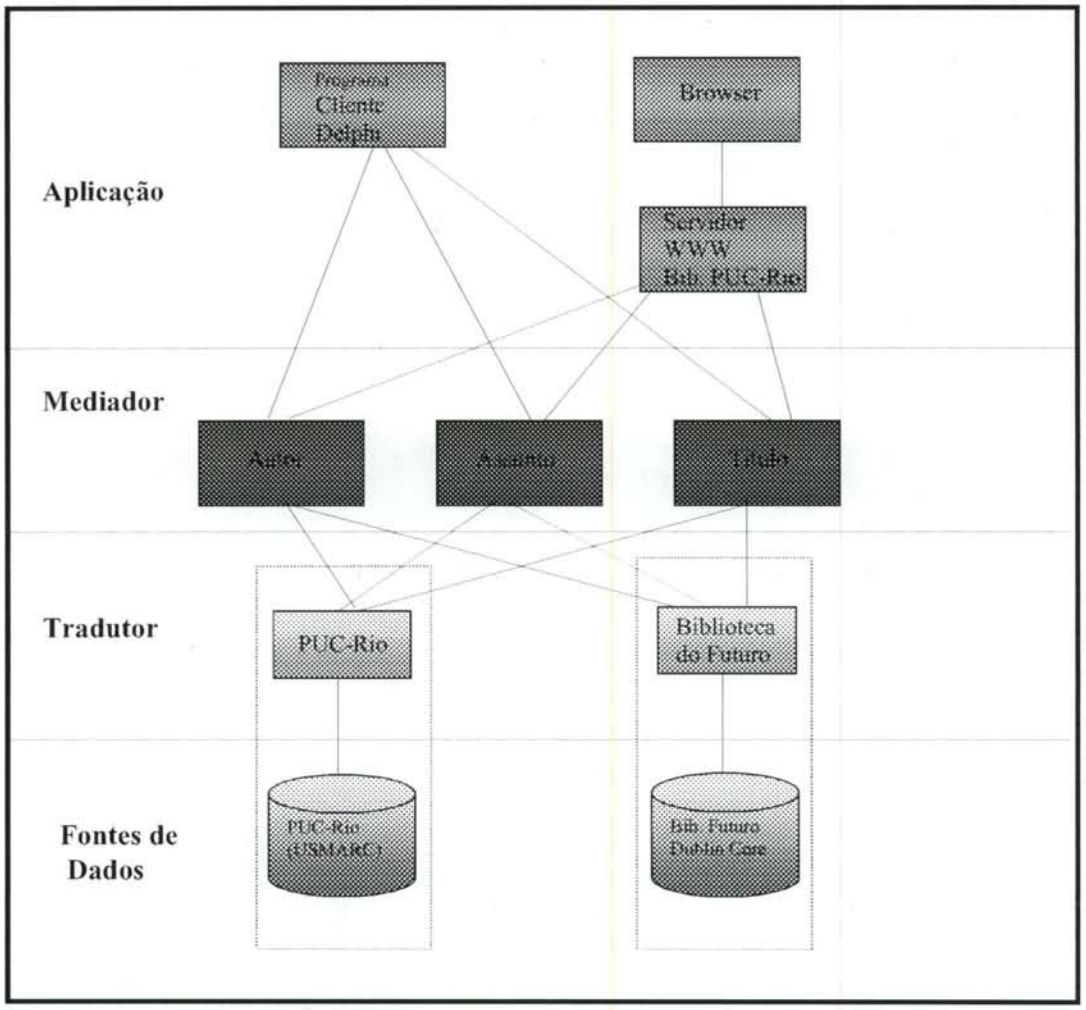

Figura 1 – Representação da Arquitetura de Integração de Dados Bibliográficos

Para que o usuário forneça os dados para sua consulta, desenvolvemos as interfaces gráficas apresentadas nas **figuras 2 e 3**, as quais possibilitam a escolha dos atributos de consulta, as fontes a serem pesquisadas, bem como o campo para os argumentos de sua pesquisa.

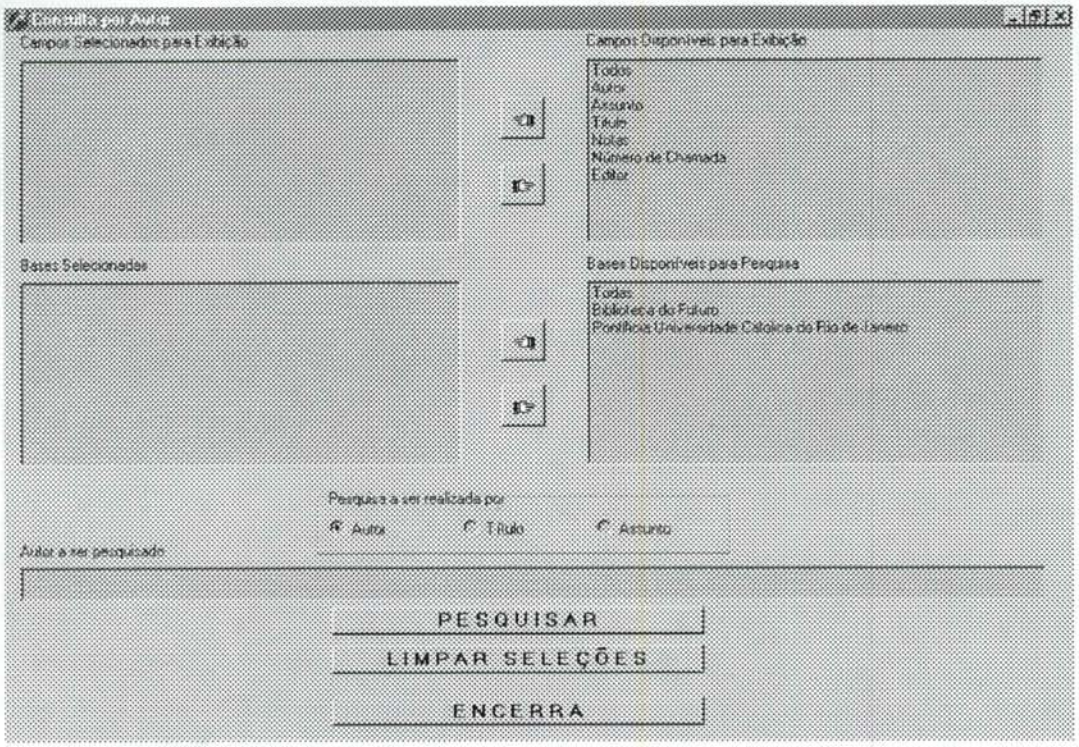

Figura 2 - Interface da Aplicação Cliente/Servidor

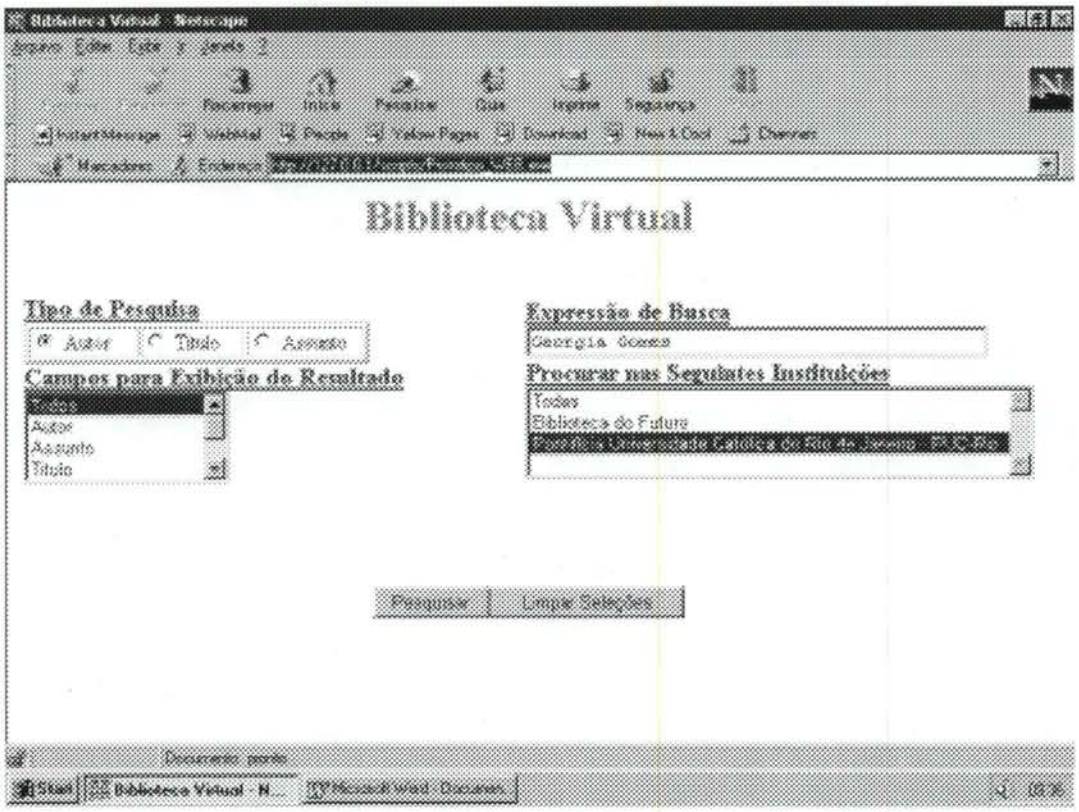

Figura 3 - Interface da Aplicação WEB

5.Conclusão

A arquitetura apresentada nestetrabalho atingiu seu objetivo, uma vez que foi possivel prover a integração entre dois sistemas de bibliotecas heterogêneos, disponibilizando aos usuários uma visão transparente e integrada de todas as bibliotecas participantes do modelo de integração, independente do modelo de dados, linguagem de consulta, sistema operacional e localização espacial.

Concluímos também que, uma arquitetura de integração deve possuir a capacidade de adaptabilidade às transformações, uma vez que o ambiente interoperável está normalmente sujeito à mudanças. A abordagem de Mediadores proposta neste trabalho propicia esse alto grau de adaptabilidade.

Referências Bibliográficas

[Barker94] Barker, Phillip. *Eletronic Libraries: visions of the future*. The Eletronic Library, Vol. 12, n.4. Aug. 1994.

[Cloyes94] Cloyes, K. *The Journey from vision to reality of a virtual library.* Special Library, Vol. 85, n. 4, Fall 1994.

[Deschamps94] Deschamps, M. C. *The Electronic library: Bielefeld Conference, 1994.* Libri, Vol. 44, n. 4, Dec. 1994.

[Gomes99] Gomes, G.R. R., *Um ambiente para integração de Dados Bibliográficos baseado em Mediadores.* Departamento de Informática da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Dissertação de Mestrado em Informática : Ciência da Computação, 1999.

[Kemp94] Kemp, A. *Eletronic Information: solving old or creating new problems.* Libri, Vol. 44, n. 4, Dec 1994.

[Piggott93] Piggott, Sylvia E.A. *The Virtual Library: almost there*. Special Libraries, Vol.84, n. 4. Fall 1993.

[Poulter94] Poulter, A. *Building a browsable virtual reality library*. Aslib Proceedings, Vol. 46, n. 6, June 1994.

[Saunders95] Saunders, L. M. *Transforming acquitions to suport virtual libraries*. Information Technology and Libraries, Vol. 14, N. 1, March 1995.

[Sumpter94] Sumpter, R., *Whitepaper on Data Management*. Lawrence Livermore National Laboratory, The IEEE Metadata Workshop, 1994.

[Weibel95] Weibel, S., Godby, J., Miller, E. (1), Daniel, R. (2), *OCLC/NCSA Metadata Workshop Report* (1) Office of Research, OCLC Online Computer Library

Center, Inc. (2) Advanced Computing Lab, Los Alamos National Laboratory. Web Site www.oclc.org:5046/oclc/usearch/conferences/metadata/ 1995.

[Yeager96] Yeager, N. J., McGrath, R. E.. *Web Server Tecnology – The Advanced Guide for World Wide Web Information Providers* , Morgan Kaufmann Publishers, Inc., 1996.