Uma Interface Web para Identificação de Equivalências em Bancos de Dados Heterogêneos

Fernando Busanello Meneghetti[1], Fabiano Gama Paes[1], Gustavo Zanini Kantorski[1]

[1]Curso de Sistemas de Informação – Universidade Luterana do Brasil (ULBRA)
Campus Santa Maria – Santa Maria – RS – Brasil

{fmeneghetti, fgpaes, gustavoz}@ cpd.ufsm.br

Abstract. *The integrated identification to distributed heterogeneous information is a major concern in large corporations and governmental institutions. The use of available Web tools turns out as an affordable, simple, and also platform and DBMS independent, alternative for schema mapping of heterogeneous databases. This paper presents a open tool for web equivalence identification in heterogeneous data sources through the Web.*

Resumo. *A identificação de informações localizadas em bancos de dados heterogêneos é fundamental para grandes empresas e instituições governamentais. O uso de ferramentas através da Web facilita e simplifica o acesso a informações localizadas em SGBDs independentes.Este artigo descreve uma ferramenta web, aberta e de baixo custo, para identificação de equivalências de objetos e atributos entre diferentes fontes.*

1. Introdução

Atualmente, muitas empresas possuem fontes de dados distribuídas setorialmente atendendo requisitos específicos, porém com informações comuns. Este fato é causado porque não ocorreu um planejamento na criação dos bancos de dados, muitas vezes pela falta de análise ou da não antecipação dos inter-relacionamentos que poderiam ocorrer com o passar do tempo. Com isso, surgem os sistemas de integração de dados, que têm como objetivo fornecer aos usuários uma interface uniforme para as diversas fontes de dados, autônomas, distribuídas e heterogêneas [Özsu 1999].

Este artigo apresenta uma ferramenta web, de código fonte aberto, cujo principal objetivo é mostrar uma interface para visualização da equivalência de objetos e atributos localizados em fontes de dados heterogêneas. A ferramenta que realiza o mapeamento de esquemas entre as fontes de dados heterogêneas está descrita nos trabalhos de [Meneghetti, Paes e Kantorski 2007a] e [Meneghetti, Paes e Kantorski 2007b].

A ferramenta proposta é parte do projeto denominado CORIDORA, desenvolvido em âmbito acadêmico na Universidade Luterana do Brasil, campus Santa Maria. O projeto CORIDORA tem como objetivo tratar inconsistências, e possíveis limpezas de dados, em bancos de dados, derivadas da representação de equivalências de um mesmo objeto do mundo real. A equivalência é realizada com o mapeamento de esquemas conceituais, identificando, consistindo e comparando divergências entre os objetos equivalentes, sem prejudicar a autonomia local das fontes de dados.

A próxima seção apresenta a metodologia utilizada para implementação da ferramenta. Na seção 3 são apresentados os casos de uso, a equivalência de objetos e atributos e um estudo de caso realizado. A seção 4 aborda alguns trabalhos relacionados e uma comparação com a ferramenta desenvolvida. Considerações finais e trabalhos futuros são apresentados na seção 5.

2. Equivalência de Objetos e Atributos

Existem inúmeras metodologias para propiciar o acesso integrado a banco de dados heterogêneos. A metodologia de mapeamento de esquemas conceituais [Ribeiro 1995] foi adotada para o desenvolvimento desta ferramenta, com o intuito de permitir que duas abstrações distintas, as quais representam um mesmo objeto do mundo real, coexistam e trabalhem harmonicamente, possibilitando assim, o acesso às diferentes facetas de um mesmo objeto como um todo.

O mapeamento de esquemas conceituais permite que os esquemas locais sejam mantidos na sua forma original permanecendo inalterada a autonomia local. A existência de um registro central (RC) tem como função a manutenção de cópias dos esquemas conceituais de exportação participantes e dos objetos correspondentes, disponíveis à comunidade [Kantorski 2000].

Neste contexto, a figura 1 ilustra o processo de mapeamento de esquemas conceituais, sendo necessário o ingresso de pelo menos duas fontes de dados na federação para tornar possível o mapeamento.

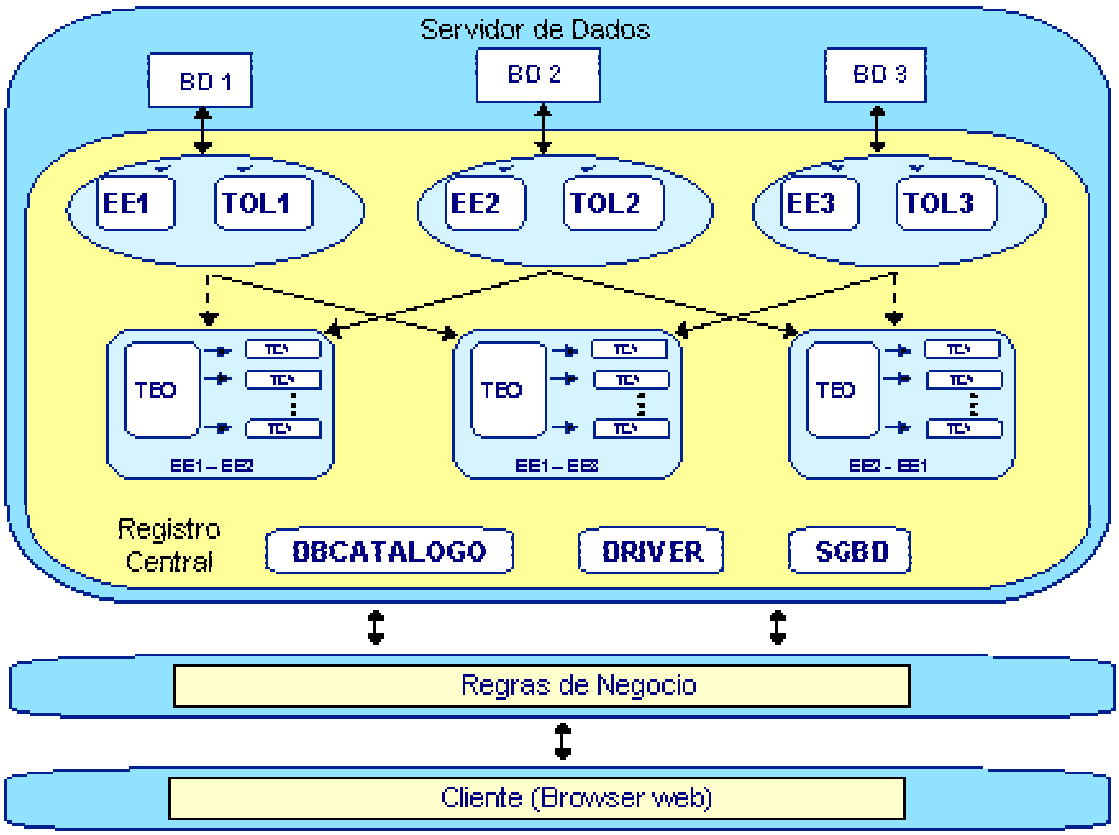

Figura 1. Metodologia de Mapeamento de Esquemas.

O processo de mapeamento de esquemas é precedido pela conversão do esquema conceitual local para o esquema de exportação (EE), através da exportação de atributos, e pela criação de uma tabela de objetos locais (TOL) contendo informações sobre o conjunto de objetos modelados localmente a partir das classes e papéis. Os papéis têm a funcionalidade de identificar uma possível função do objeto no mundo real [Ribeiro 1995]. Uma característica importante na exportação para o EE é que deve ser definida a propriedade que identifica unicamente (*unique*) aquele objeto na federação.

Desta forma, é no registro central que está armazenada uma cópia das tabelas de Esquemas de Exportação (EE) e da tabela de Objetos Locais (TOL) do novo membro da federação. O esquema local passa, então, a ter acesso ao conjunto de EEs e TOLs disponíveis no RC, podendo selecionar aqueles esquemas externos aos quais deseja mapear.

Assim são geradas a Tabela de Equivalência de Objetos(TEO) e a Tabela de Equivalência de Atributos(TEA). Nestas tabelas estão registradas as equivalências e divergências de representação dos diversos objetos, de forma a tornar possível o acesso integrado ao conjunto total de instâncias correspondentes a uma mesma entidade do mundo real, distribuídas através dos bancos de dados da federação [Kantorski 2000].

3. Ferramenta de Identificação de Equivalências

Logado no ambiente o usuário possui acesso liberado às funcionalidades disponibilizadas pela ferramenta, tais como a manutenção das fontes de dados pertencentes a Federação, exportação dos dados disponibilizando a comunidade, realização da equivalência de objetos e atributos [Meneghetti, Paes e Kantorski 2007a].

Após existirem pelo menos duas fontes de dados participando da federação, possuindo objetos e atributos presentes no esquema de exportação, é possível realizar a equivalência entre os objetos.

3.1 Casos de Uso

Os casos de uso que representam os tratamentos das equivalências dividem-se em dois: equivalências de objetos e equivalências de atributos.

As equivalências de objetos são propostas pelo sistema Coridora. O usuário tem acesso através do menu, às equivalências de objetos, onde a ferramenta lista as possíveis equivalências baseadas em regras pré-definidas e possibilita ao usuário adicionar ou excluir uma equivalência proposta anteriormente.

A Figura 2 ilustra o processo de adicionar ou remover equivalências de Objetos.

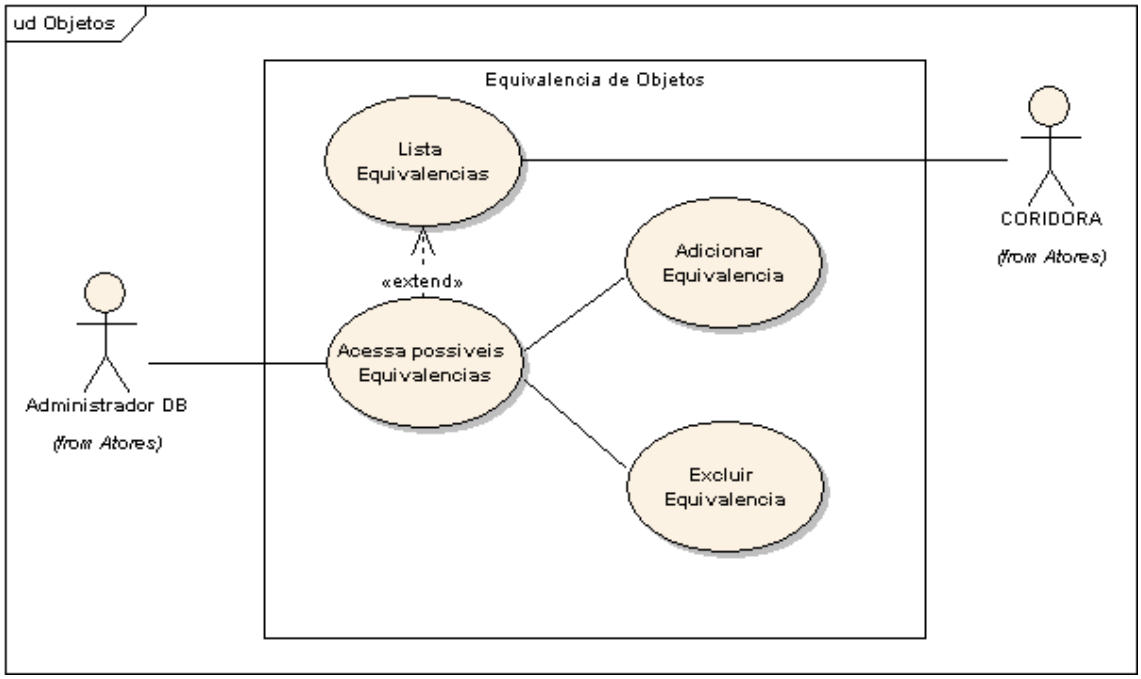

Figura 2. Caso de Uso - Equivalência de Objetos

As equivalências de atributos estão baseadas nos objetos que são equivalentes, ou seja, após o usuário acessar através a interface de equivalência dos atributos, são listados pelo sistema todos objetos que o usuário confirmou como equivalentes, com a possibilidade de verificar e alterar a equivalência entre seus atributos.A figura 3 exemplifica o processo de identificação de equivalência de atributos.

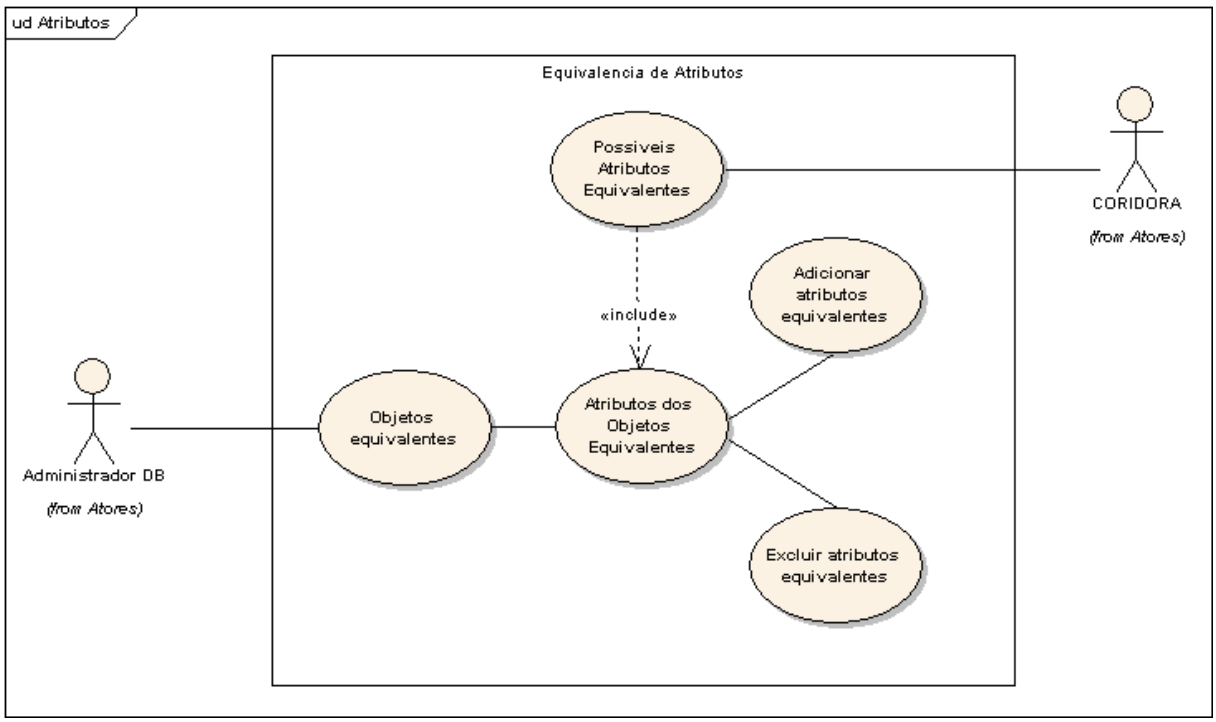

Figura 3. Caso de Uso - Equivalência de Atributos

Ao realizar a seleção para verificar as equivalências referentes aos atributos, é aberta uma janela *popup* a qual possibilita ao usuário visualizar as equivalências proposta pelo sistema, baseando-se no princípio que nomes de atributos iguais são equivalentes. Desta forma possibilita o usuário aceitar as equivalências propostas pela ferramenta, excluí-las ou ainda adicionar uma nova equivalência que não foi identificada.Todas equivalências que forem identificadas e aceitas pelo usuário, ou ainda as propostas pelo próprio usuário são registradas na tabela de equivalência de atributos (TEA).

3.2 A Interface da Ferramenta

As equivalências dos objetos seguem uma estratégia binária [Batini 1986], onde pares de esquemas são comparados e as equivalências são propostas pelo ambiente seguindo algumas regras pré-definidas listadas a seguir e que podem ser visualizadas na figura 4:

- Regra 1: dois objetos com o mesmo nome na TOL, em fontes diferentes, são equivalentes;
- Regra 2: nomes de atributos iguais no EE e ambos com o mesmo *unique* são equivalentes. Nomes de atributos de uma fonte igual a nomes de objetos de outra fonte são equivalentes. E, nome de objeto de uma fonte seja igual ao nome de atributo de outra fonte são equivalentes;
- Regra 3: papéis – o objeto não está na TEO e o papel está na TEO ou o objeto está na TEO e o papel não está na TEO ou o papel está na TOL e o papel não está na TEO e o objeto não está na TEO são equivalentes.

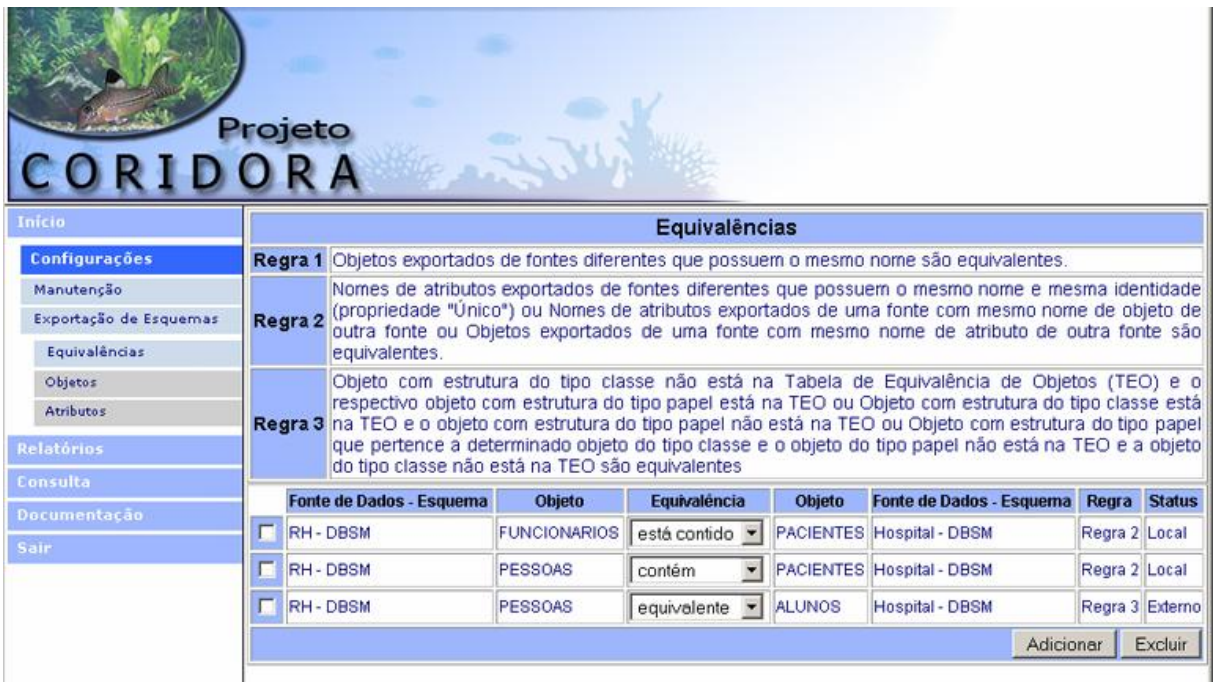

Figura 4. Interface de Identificação de Equivalência de Objetos.

Para cada par de equivalências de objetos deve-se determinar as possíveis equivalências entre suas propriedades. Assim, a ferramenta desenvolvida possui uma interface que apresenta as equivalências de atributos destacando-se a identificação automática das diversas equivalências. Nesta mesma interface, são apresentados os

atributos propostos pelo ambiente e aceitos pelo usuário com sua célula na cor vermelha. Já as equivalências propostas pela ferramenta e não aceitas pelo usuário encontram-se na cor amarela. Finalmente, as equivalências de atributos incluídas pelo usuário não identificadas pela ferramenta são apresentadas na cor verde. A interface da ferramenta pode ser visualizada na figura 5.

Essas interfaces tornam a ferramenta mais intuitiva, pois apresenta ao usuário, as possíveis equivalências de objetos já adicionadas, para através dela chegar a interface de equivalências de atributos.

Figura 5. Interface de Identificação de Atributos Equivalentes

3.3 Estudo de Caso

Para demonstrar as funcionalidades referentes à identificação das equivalências entre objetos e atributos, um estudo de caso foi realizado na Universidade Federal de Santa Maria (UFSM), mais especificamente utilizando o banco de dados do Hospital Universitário de Santa Maria (HUSM), do setor de recursos humanos e os dados do Restaurante Universitário (RU).

São apresentados três modelos, enfocando universos diferentes, mas com características de aplicação semelhantes. A primeira aplicação modelada refere-se ao ambiente clínico, mais especificamente ao tratamento de pacientes internados e realização de exames. A segunda aplicação refere-se ao modelo administrativo, no que diz respeito a área de recursos humanos da Universidade. O terceiro enfoca o ambiente de restaurantes universitários da instituição. A relação entre esses três ambientes está vinculada através dos objetos locais de cada um. Com o intuito de facilitar o mapeamento dos esquemas conceituais, os esquemas de exportação dos ambientes, são propostos nas tabelas 1, 2 e 3, respectivamente. Maiores detalhes para geração dos

esquemas de exportação e tabelas de objetos locais podem ser visualizados em [Meneghetti, Paes e Kantorski 2007a] e [Meneghetti, Paes e Kantorski 2007b].

Tabela 1. Informações Exportadas do Hospital

Classe	Papel	Atributos
Pacientes		dt_nascimento, fator_rh, grupo_sanguineo, id_pessoa, nome_mae, nome_pai, nome_paciente, num_prontuario, sexo
Alunos		dt_nascimento, id_aluno, id_pessoa, nome_mae, nome_pai, sexo

Tabela 2. Informações exportadas do Restaurante Universitario

Classe	Papel	Atributos
Usuários		nome_usuario

Tabela 3. Informações Exportadas do RH

Classe	Papel	Atributos
Funcionários		fator_rh, id_pessoa, nome_funcionario, nome_mae, nome_pai, sexo, tipo_sanguineo
Pessoas		id_pessoa, nome_pessoa
	Alunos	dt_nascimento, id_aluno, id_pessoa, nome_pai, nome_mae, sexo

O processo de identificação das equivalências acontece em dois momentos distintos, no primeiro, são identificados os objetos equivalentes e no segundo, a equivalência de seus atributos.

A identificação dos objetos equivalentes dá-se de acordo com regras pré-estabelecidas e descritas na seção 2. No final da identificação dos objetos equivalentes chegamos ao resultado representado na tabela 4.

Tabela 4. Identificação de Objetos Equivalentes

Objetos	Ambiente/Objeto	Equivalência	Ambiente/Objeto	Regra
01	RH/Funcionarios	Está Contido	Hosp/Pacientes	Regra 2
02	RH/Pessoas	Contém	Hosp/Alunos	Regra 3

A identificação dos atributos equivalentes ocorre de acordo com cada par de objetos equivalentes relacionados na TEO. Após o ambiente montar a interface mostrando os atributos específicos de cada objeto, são listadas possíveis equivalências baseando-se nas regras estabelecidas, conforme apresentado na tabela 5.

Tabela 5. Equivalência de Atributos

Funcionário / **Paciente**	fator_ rh	id_ pessoa	nome_ funcionario	nome_ mae	nome_ pai	sexo	tipo_ sanguineo
dt_ nascimento							
fator_rh	equiv.						
grupo_ sanguineo							equiv.
id_pessoa		equiv.					
nome_mae				equiv.			
nome_ paciente			equiv.				
nome_pai					equiv.		
num_ prontuario							
Sexo						equiv.	

Ao final do estudo de caso chegou-se ao resultado esperado, onde foi possível realizar todo o mapeamento dos esquemas e identificação das equivalências entre objetos e atributos, provando a eficiência da ferramenta.

4. Trabalhos Relacionados

Na literatura existem diversas propostas de sistemas voltados para o problema de integração de dados heterogêneos. A grande maioria desses sistemas possibilita apenas trabalhar em cima de consultas, não se preocupando em como realizar a integração das fontes heterogêneas e a identificação de equivalências entre objetos e atributos. Dentre as ferramentas estudadas destacam-se: DataXTurbo [Carestiato 2006], FIE [Lima 1997], FIEC [Frederes 1999].

A ferramenta para integração de esquemas conceituais proposta por [Lima 1997] foi planejada para operar de acordo com as especificações da arquitetura CORBA, possibilitando assim a resolução de heterogeneidades no nível de plataforma. É composta por dois módulos: a) módulo servidor, responsável pelo armazenamento e recuperação das tabelas de equivalências, esquemas de exportação e tabela de objetos locais. É composto de um adaptador de objetos, que possibilita a conexão de banco de dados POStgres 95 com as implementações de servidores de objetos da ferramenta e b) módulo cliente, responsável pelo mapeamento das equivalências, detecção dos conflitos e geração da interfaces HTML para comunicação com o usuário.

A ferramenta DataXTurbo foi desenvolvida através de drivers JDBC para acesso aos dados. A configuração das diversas visões do usuário é realizada através de arquivos XML que permitem ao desenvolvedor alterar as consultas geradas pela aplicação, alterando assim o comportamento da aplicação de forma simples sem a necessidade de recompilação [CARESTIASO, 2006].

A ferramenta proposta por [FREDERES, 1999] possibilita a conversão de esquemas locais relacionais para o esquema de exportação, representado em um modelo canônico, com a criação em paralelo de uma tabela de objetos locais. Com esta

ferramenta de conversão disponibilizada, é possível realizar conversões de esquemas de bancos de dados quaisquer para esquemas orientados a objetos, estendendo sua aplicação a metodologias que necessitem da conversão dos esquemas conceituais locais para um modelo comum. O único requisito apresentado pela ferramenta é de um arquivo SQL representado as estruturas do esquema original. A ferramenta apresenta uma interface não muito clara na identificação das equivalências de objetos e atributos.

Em relação a outras ferramentas encontradas na literatura, a ferramenta desenvolvida contribui através da independência de plataforma, por ser *open source* e pela interface amigável que foi desenvolvida permitindo que os usuários não tenham conhecimentos em programação e configuração de arquivos XML. Na tabela 6 pode ser visualizado um comparativo entre as ferramentas estudadas e a desenvolvida.

Tabela 6. Ferramentas para Acesso a Bancos de Dados Heterogêneos

Indicadores	**DataXTurbo**	**FIE**	**FIEC**	**Coridora**
Independência de plataforma	Total	Parcial	Parcial	Total
Suporte a acesso via Web	Sim	Não	Não	Sim
Licença	Livre	Proprietária	Proprietária	Open Source
Recursos para identificar as equivalências	Conhecimentos Em XML	Conhecimentos em SQL	Conhecimentos em SQL	Nenhum
Tecnologias utilizadas na implementação	XML	-	-	JSP

5. Considerações Finais

Este trabalho apresenta uma ferramenta web, de código fonte aberto, que realiza a identificação de equivalências de objetos e atributos entre fontes de dados heterogêneas, com o intuito de não realizar alterações nas estruturas e dados que se encontram localmente.

A partir da metodologia proposta por [Ribeiro 1995] e da ferramenta de mapeamento de esquemas desenvolvida por [Meneghetti, Paes e Kantorski 2007a] e [Meneghetti, Paes e Kantorski 2007b] que realiza o mapeamento dos esquemas conceituais onde os modelos locais são convertidos e passam a ser um único modelo, foi possível desenvolver uma ferramenta que identifica as equivalências entre diversas fontes de dados que se encontram distribuídas logicamente e fisicamente em ambientes heterogêneos.

A ferramenta desenvolvida tem a característica de ser dinâmica, intuitiva e de fácil uso não necessitando que os usuários possuam conhecimentos avançados relacionados a bancos de dados e ambientes heterogêneos. As informações que os usuários devem ter para realizar as equivalências entre objetos e atributos são aquelas referentes ao modelo conceitual de dados de cada fonte participante. Desta forma podem ser acessados os diversos objetos, atributos e possíveis equivalências entre as diversas fontes. A ferramenta está disponível em http://portal.ufsm.br/coridora/.

Os resultados alcançados com este artigo abrem a perspectiva de trabalhos futuros. Uma possibilidade é a implementação de algoritmos de consulta que possibilitem acessar os dados contidos nessas fontes de dados heterogêneas de uma forma uniforme e complementar. Além das ferramentas de consulta, a implementação de algoritmos de similaridade também pode ser incluída como trabalhos futuros, integrando e identificando divergências entre as diversas fontes participantes.

Na versão projetada e desenvolvida para este trabalho, as equivalências dos objetos somente são listadas e podem ser adicionadas conforme regras pré-definidas pelo ambiente, não sendo possível criar novas equivalências. No estudo de caso realizado verificou-se que as informações exportadas da fonte do restaurante não atenderam nenhuma das regras, o que dificulta o mapeamento. Fica como uma proposta futura a ampliação desta funcionalidade para atender ao requisito de adição de novas equivalências entre objetos, não levando em conta as regras estabelecidas, permitindo a possibilidade de selecionar quais objetos dentre os exportados são equivalentes.

Referências

Batini, C., Lenzerini, M., Navathe, S. (1986) "*A Comparative Analysis of Methodologies for Database Schema Integration."* ACM Computing Surveys, New York, v.18, n.4, p.323-364, Dez.

Carestiato, B. (2006) "Ferramenta para Integracao de Banco de Dados Heterogêneos." Monografia. Rio de Janeiro: PUC-RIO.

Frederes, S., Ribeiro, C.H.F.P. Pallazo, J. (1999) "Ferramenta de Apoio a Conversão de Esquemas Conceituais Heterogêneos" – CPGCC, Porto Alegre, UFRGS.

Lima, J.C.D., Ribeiro, C. H. F. P., Palazzo, J. (1997) "Acesso Integrado a Bancos de Dados Distribuídos Heterogêneos utilizando CORBA**."** In: SBBD, 1997. UFRGS.

Kantorski, G. Z. ; Ribeiro, C. H. F. (2000) "*Heterogeneous Database Interoperability using the WWW"*. In: Simpósio Brasileiro de Banco de Dados, João Pessoa. XV Simpósio Brasileiro de Banco de Dados, 2000. p. 79-88.

Meneghetti, F. B., Paes, F. G., Kantorski, G. Z. (2007a) " CORIDORA *Mapping:* Uma Ferramenta Web para Mapeamento de Equivalências Semânticas em Bancos de Dados Heterogêneos." In: Simpósio de Informática, 2007, Uruguaiana – RS. XII Simpósio de Informática, Nov.

Meneghetti, F. B., Paes, F. G., Kantorski, G. Z. (2007b) "Ferramenta CORIDORA *Mapping* para Mapeamento de Esquemas em Bancos de Dados Heterogêneos"**.** In: Seminário de Informática, Torres – RS. VII Seminário de Informática, Nov.

Ribeiro, Cora Helena Francisconi Pinto. (1995) "Banco de Dados Heterogêneos: Mapeamento dos Esquemas Conceituais em um Modelo Orientado a Objetos**" (**CPGCC**).** Porto Alegre: UFRGS, 165p.

Özsu, M.T., Valduriez, P., (1999) "*Principles of distributed database systems*." 3.ed. Prentice-Hall, 1999.