



**UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA**

Relatórios Técnicos
do Departamento de Informática Aplicada
da UNIRIO
n°0001/2009

Integração Semântica de Modelos

**Jairo Souza
Thiago Andrade
Martin Pereira
Mauro Lopes
Sean Siqueira
Fernanda Baião
Claudia Cappelli**

Departamento de Informática Aplicada

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
Av. Pasteur, 458, Urca - CEP 22290-240
RIO DE JANEIRO – BRASIL

Projeto de Pesquisa

Grupo de Pesquisa Participante



Patrocínio



PETROBRAS

Integração semântica de modelos *

Jairo Souza^{1,2}, Thiago Andrade², Martin Pereira², Mauro Lopes², Sean Siqueira^{2,3}, Fernanda Baião^{2,3}, Claudia Cappelli^{1,2}

¹Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO)

²Núcleo de Pesquisa e Prática em Tecnologia (NP2Tec) – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

³ Departamento de Informática Aplicada (DIA) – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

jsouza@inf.puc-rio.br, thiago.andrade@uniriotec.br,
martin.pereira@uniriotec.br, mauro.lopes@uniriotec.br, sean@uniriotec.br,
fernanda.baiao@uniriotec.br, claudia.cappelli@uniriotec.br

Abstract. For an efficient data administration (DA), the DA team needs to access rich representations of the concepts of the domain of discourse. This is possible through business process models and domain ontologies, for instance. This work proposes an approach for semantic annotation of process models, logical data models, and database schemas, based on a domain ontology. The semantic annotations also made it possible for creating an automatic interface among some common tools used for data and process modeling purposes.

Keywords: Semantic integration; knowledge management; ontology.

Resumo. Para a realização eficiente das suas atividades, é essencial que a equipe de administração de dados (ADs) tenha acesso aos artefatos que permitam conhecer os conceitos sobre o domínio dos dados que estão sob sua responsabilidade. Dentre os artefatos disponíveis estão, por exemplo, modelos de processo de negócio e as ontologias de domínio. Este trabalho propõe uma abordagem de enriquecimento semântico de modelos de processos, modelos lógicos de dados e esquemas de banco de dados com uma ontologia de domínio através de ferramentas conhecidas do mercado.

Palavras-chave: Integração semântica; gestão do conhecimento; ontologia.

* Trabalho patrocinado pela Petrobras.

Sumário

1	Introdução	6
1.1	Motivação	6
1.2	Estrutura do Relatório	7
2	Integração de modelos de processos e ontologias	7
2.1	Adaptação do filtro	8
2.2	Script	11
2.2.1	Importação do script	11
2.2.2	Conteúdo do script	12
3	Integração de modelos lógicos de dados e ontologias	13
3.1	Formas possíveis de integração do Erwin com Protégé	13
3.1.1	Macro	13
3.1.2	UDP	14
3.1.3	API	14
3.1.4	XML	15
3.2	Proposta de utilização do Erwin para integração com Protégé	15
3.2.1	Criação De UDP	16
3.2.2	Exportar modelo lógico para o formato XML	16
3.2.3	Interface que transforma o XML exportado	17
4	Integração de base de dados com ontologias	19
4.1	Detalhamento da ferramenta Oracle 11g	20
4.2	Comunicação com o Protégé	20
5	Considerações sobre as soluções propostas de integração	23
5.1	Falta de um repositório de ontologias	23
5.2	Definições de infra-estrutura necessárias à execução da solução proposta	24
5.3	Limitação de licenças para o ARIS	24
5.4	Limitações da comunicação com o Oracle	25
6	Conclusões	26

1 Introdução

Este documento tem como objetivo apresentar o estudo para uma abordagem de integração de modelos através de anotações semânticas, como parte do Projeto de Pesquisa em Integração de Modelos com Ontologias, dentro do Termo de Cooperação em Modelagem de Processos, estabelecido entre a TIC-E&P/GDIEP e o NP2Tec/UNIRIO.

1.1 Motivação

Para a realização eficiente das suas atividades, uma equipe de administração de dados (ADs) necessita ter conhecimento sobre o domínio dos dados que estão sob sua responsabilidade. Tal conhecimento permite avaliar a modelagem de dados existente, propor alterações no esquema de dados que estejam de acordo com a utilização dos dados pelas atividades do negócio, e a interpretação comum dos dados armazenados quanto ao seu significado entre a equipe de ADs e os usuários dos dados. É essencial, portanto, que os ADs tenham acesso a artefatos que os permitam conhecer os conceitos do domínio que deram origem aos modelos dos dados sob sua responsabilidade. Dentre os artefatos possíveis estão, por exemplo, modelos de processo de negócio e as ontologias de domínio.

A modelagem de processos de negócio necessariamente faz uso de um conjunto de conceitos do domínio para descrever as diversas visões que compõem o modelo do negócio. Em muitos casos, no entanto, tais conceitos não possuem significado comum, consensual e compartilhado entre todos os envolvidos com o processo, o que certamente compromete a interpretação correta do modelo de negócio representado, além de ser fonte de diversos problemas de modelagem e integração de dados. Desta forma, a integração de uma ferramenta de modelagem de processos e uma ferramenta de engenharia de ontologias visa permitir ao usuário utilizar os dois ambientes simultaneamente, podendo assim esclarecer quaisquer dúvidas quanto à semântica dos conceitos utilizados no modelo de processos.

Ontologias são estruturas usadas para representar conhecimento de domínio e, principalmente, conhecimento consensual. O modelo de um banco de dados, por sua vez, não possui ligação com uma representação de mais alto nível que representem os conceitos que tabelas ou atributos armazenam e, dessa forma, é fonte de diversos problemas de modelagem durante a sua manutenção.

Portanto, a finalidade do projeto é permitir aos administradores de dados, através de uma solução computacional viável, um maior relacionamento entre os conceitos que estão representados nas ontologias de domínio e (1) os processos de negócio modelados, (2) os modelos lógicos de dados e (3) o esquema físico da base de dados.

Para a manipulação dos modelos de processos de negócio foi utilizada a ferramenta Aris, adotada na Petrobras para modelagem de processos de negócio. A suíte ARIS [IDS-SCHEER, 2008] oferece suporte à criação, manutenção e exibição de modelos de processos, assim como a emissão de relatórios personalizados dos modelos de processos existentes.

Para a manipulação dos modelos lógicos foi utilizada a ferramenta Erwin [CA, 2008a], adotada na Petrobras para modelagem lógica e física de dados. O Erwin oferece suporte à criação, manutenção e exibição de modelos lógico e físicos. Foi utilizado também o Oracle 11g [ORACLE, 2008a] como sistema gerenciador de banco de dados.

No uso da ontologia de domínio utilizou-se a ferramenta Protégé [PROTÉGÉ, 2008], um software livre em Java que permite a visualização, manutenção e edição de ontologias, com suporte a importação e exportação de arquivos que contenham uma ontologia de domínio. Protégé é uma das ferramentas de edição de ontologias mais utilizadas, tanto no meio acadêmico quanto comercial [CARDOSO, 2007].

Como parte de tal solução computacional, foram estudadas alternativas de interoperabilidade entre as ferramentas (1) Aris e Protégé, (2) Erwin e Protégé e (3) Oracle e Protégé, incluindo as implementações necessárias para realizar a comunicação entre essas ferramentas.

1.2 Estrutura do Relatório

O capítulo 1 desse relatório apresentou uma introdução e motivação para o problema que pretendemos trabalhar.

O capítulo 2 demonstra as adaptações realizadas na ferramenta Aris para comunicação com o Protégé. As adaptações realizadas com o mesmo objetivo na ferramenta Erwin são apresentadas no capítulo 3.

O capítulo 4 apresenta o estudo das funcionalidades semânticas existentes na versão 11g da ferramenta Oracle e finaliza com a avaliação da sua comunicabilidade com a ferramenta Protégé.

O capítulo 5 apresenta as limitações das soluções implementadas na ferramenta Aris e Erwin, além das limitações semânticas do Oracle.

Em seguida são apresentadas as conclusões da proposta no capítulo 6, e as referências bibliográficas.

2 Integração de modelos de processos e ontologias

Como forma de enriquecer semanticamente os elementos do modelo de processo para permitir um maior relacionamento entre os processos de negócio modelados e os conceitos que estão representados nas ontologias de domínio, decidiu-se realizar anotações semânticas nos elementos do modelo de processo. Chama-se anotação semântica, tecnicamente, a associação de uma URI que relaciona um artefato (no nosso caso, elementos do modelo; sendo a URI uma de suas propriedades) a um recurso de rede. Estes recursos de rede são metadados compreensíveis pela máquina, geralmente ontologias descritas em linguagem OWL ou RDF [OREN *et al*, 2006].

A ferramenta ARIS não oferece suporte para anotações semânticas com ontologias. Assim, foi necessário realizar adaptações na ferramenta que permitisse anotar elementos do modelo, ou seja, associar as URIs dos conceitos como propriedades dos elementos do modelo ARIS. Para facilitar a manipulação dessas anotações e sua visualização, foi utilizada uma versão customizada para o projeto da ferramenta Protégé, a qual não possui funcionalidades de edição, impedindo que pessoas não autorizadas alterem a ontologia visualizada. Assim, foi necessário viabilizar uma forma de integração entre a ferramenta ARIS e a ferramenta Protégé.

As seções seguintes detalham as modificações realizadas na ferramenta ARIS para viabilizar a funcionalidade de anotações semânticas na ferramenta.

2.1 Adaptação do filtro

Primeiramente é necessário criar dois atributos para os elementos ARIS. Esses atributos conterão as referências para a ontologia e o conceito correspondente ao elemento ARIS. A criação de um novo grupo se justifica para melhor organização destes atributos, destacando o diferencial que há no tratamento das informações presentes nesse grupo. Para tal, no módulo de administração do Aris (acessível através do menu View > Modules > Administration), em Attribute type groups (que pela árvore de navegação fica em Configuration > Method), cria-se um grupo onde ficarão reunidas as 2 novas propriedades semânticas dos objetos do modelo de processos.

Utilizando um atributo livre de grupo ainda disponível pela aba “Contents” (que em nosso caso particular de estudo foi utilizado o de nome “Free attribute type group 1”, possuindo o respectivo API name AT_FREE_ATG_1), clicando com o botão direito e em seguida indo em properties do atributo livre de grupo, na janela de propriedades que surgirá, definir para o campo “Language” tanto quando selecionado como “Portuguese (Brazil)” e “English (United States)”, preencher no campo Name em ambos os casos como “Grupo semântico”. Após as definições nos campos, confirmar as novas informações clicando no botão “Ok”. Na Figura 1, a conclusão das novas informações do grupo.

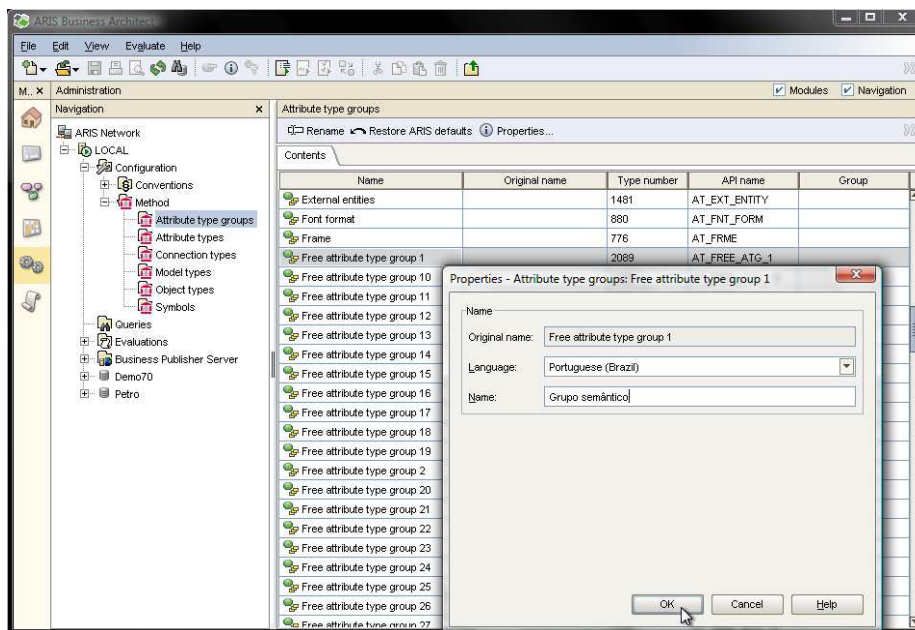


Figura 1 - Configuração do atributo “Grupo Semântico”

Após confirmação, aparecerá listada na aba “Contents” a substituição de “Free attribute type group 1” por “Grupo semântico”, conforme mostra a próxima figura.

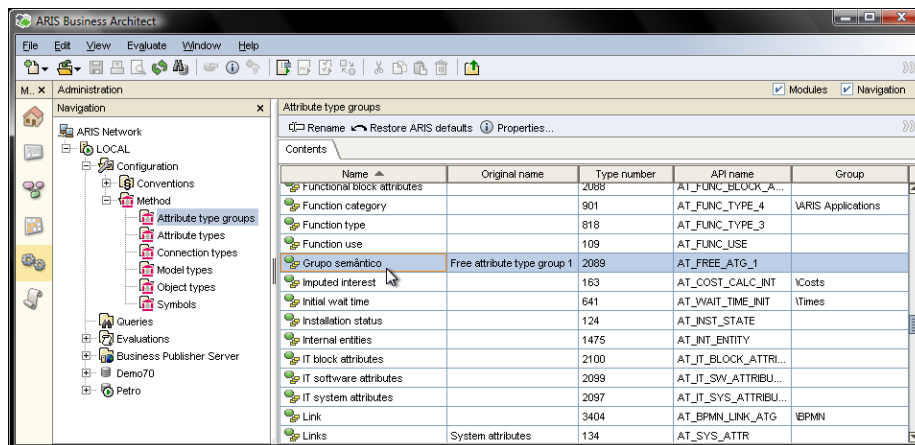


Figura 2 - Grupo semântico criado em Attribute type groups

Criado o grupo semântico, em Attribute types (na árvore de navegação fica em Configuration > Method) criam-se as 2 propriedades semânticas dos objetos do modelo de processos (“Anotação Semântica” e “Namespace da Ontologia”). Essas novas propriedades associam o elemento do modelo de processos ao conceito correspondente na ontologia.

Quando escolhidos os 2 atributos de texto ainda disponíveis dentre os listados na aba Contents, preencher no campo Name “Anotação semântica” para um atributo escolhido (que em nosso caso de estudo foi o de API name “AT_USER_ATTR2”) e “Namespace da ontologia” para o outro atributo escolhido (que nesse caso foi o que API name “AT_UA_TXT_20”).

Ainda na janela “properties”, ao clicar no botão “Change...”, surge uma janela para definir em qual grupo deve pertencer o atributo. Seleciona-se então o “Grupo semântico” presente na árvore de navegação, e em seguida para confirmar as novas informações clica-se nos botões “Ok” das janelas.

Na figura a seguir, o momento de conclusão das novas informações de um atributo.

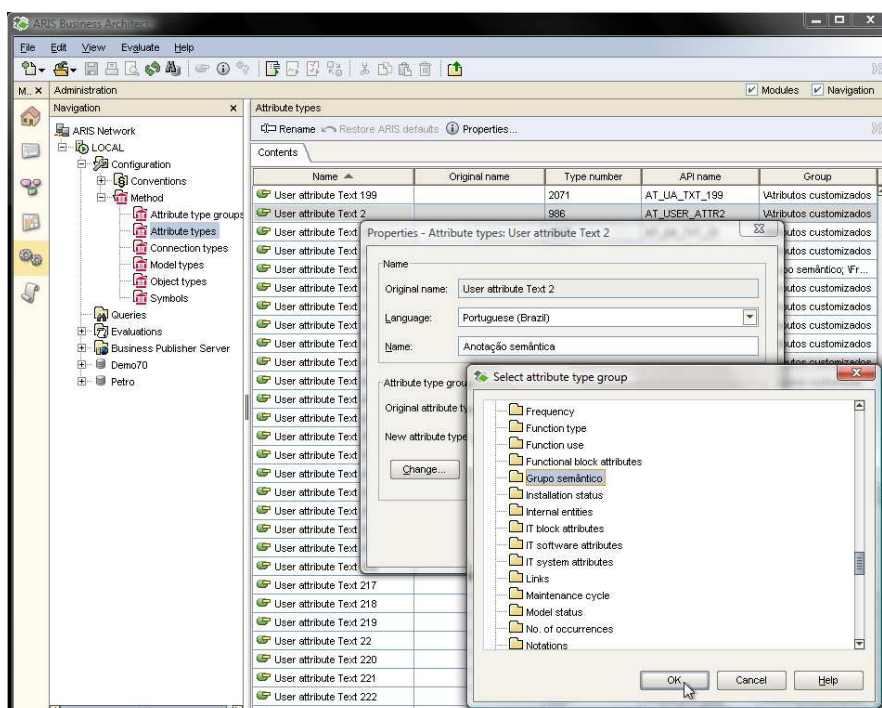


Figura 3 - Configuração do atributo livre de texto

Após as confirmações, aparecerão listadas na aba “Contents” a substituição de “User attribute Text 2” e “User attribute Text 20” por “Anotação semântica” e “Namespace da ontologia”, respectivamente, conforme na figura seguinte.

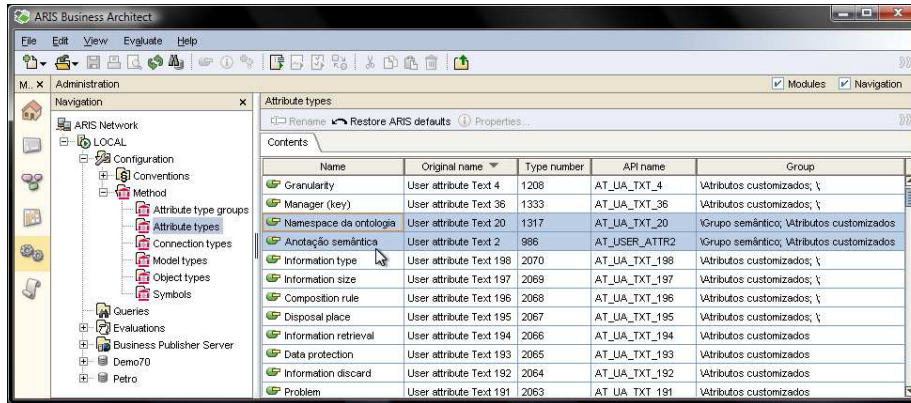


Figura 4 - Propriedades criadas em Attribute types

Com as propriedades criadas e associadas ao grupo, é preciso agora acrescentá-las no filtro metodológico para que as novas propriedades estejam disponíveis durante a modelagem de processos. Ainda no módulo de administração do Aris, em “Filter” (pela árvore de navegação fica em Configuration > Conventions) localize o “Filtro Gestao Processos - Corporativo Petrobras”, na aba “Contents”, e clicando com o botão direito e em seguida indo em “Edit...” surgirá a janela “Filter Wizard”.

Clicando em “Next” até o passo “10. Select object attributes”, selecione todos os itens em “Object types” e em “Attribute type group” selecione o “Grupo semântico”. Com isso, seleciona-se “Anotação semântica” e “Namespace da ontologia” em “Attribute types”, concluindo as alterações com o clique em “Finish”, conforme na Figura 5.

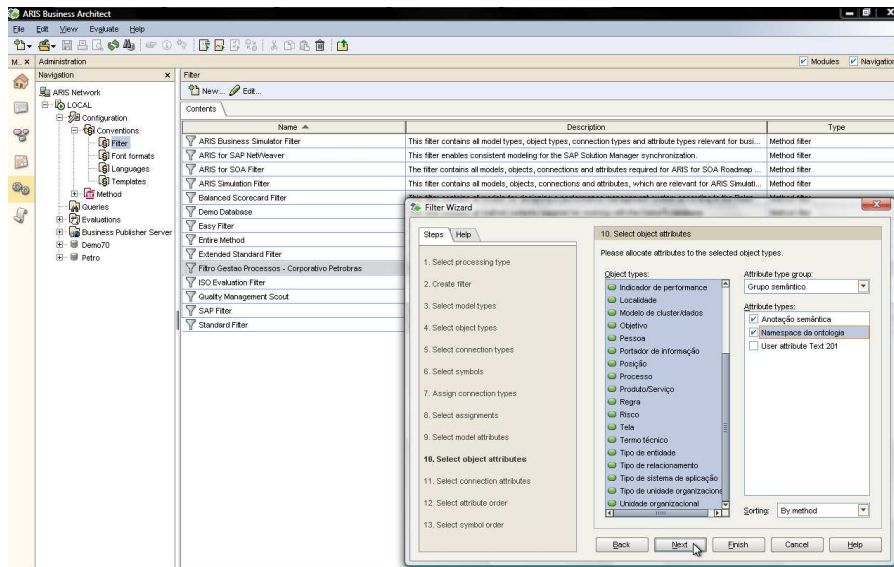


Figura 5 - Associação dos atributos ao filtro metodológico

2.2 Script

A integração do Aris com o Protégé ocorre através de um script, codificado em linguagem JavaScript, como forma de interface entre os softwares.

Neste script, seleciona-se todos os conceitos envolvidos em uma determinada atividade do processo, acionando então o Protégé, que exibe a ontologia com os respectivos termos.

A seguir, é detalhado o script criado para a integração do Aris com o Protégé.

2.2.1 Importação do script

No módulo “Script Editor” (pelo menu View > Modules > Script Editor), demonstrado na Figura 6, na árvore de navegação, clica-se com o botão direito em “Reports” e na seqüência “New” > “Category” cria-se a nova categoria, no caso nomeada como “Semântica”, onde ficará o script.

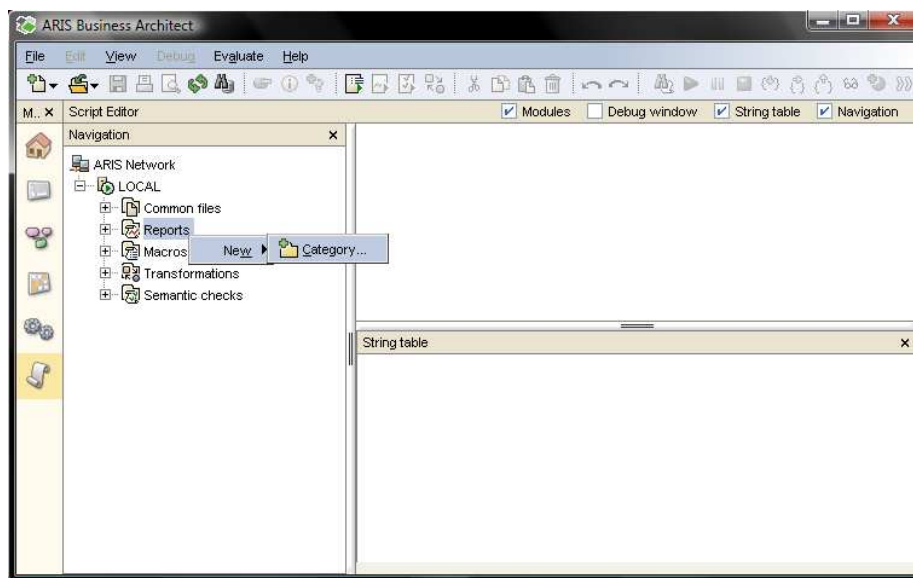


Figura 6 - Criação de nova categoria

Para a importação, na árvore de navegação clique com o botão direito em “Semântica”, e na seqüência em “Import...”, demonstrado na figura seguinte, e na tela de navegação que surgir, abrir o arquivo que contém o script desejado.

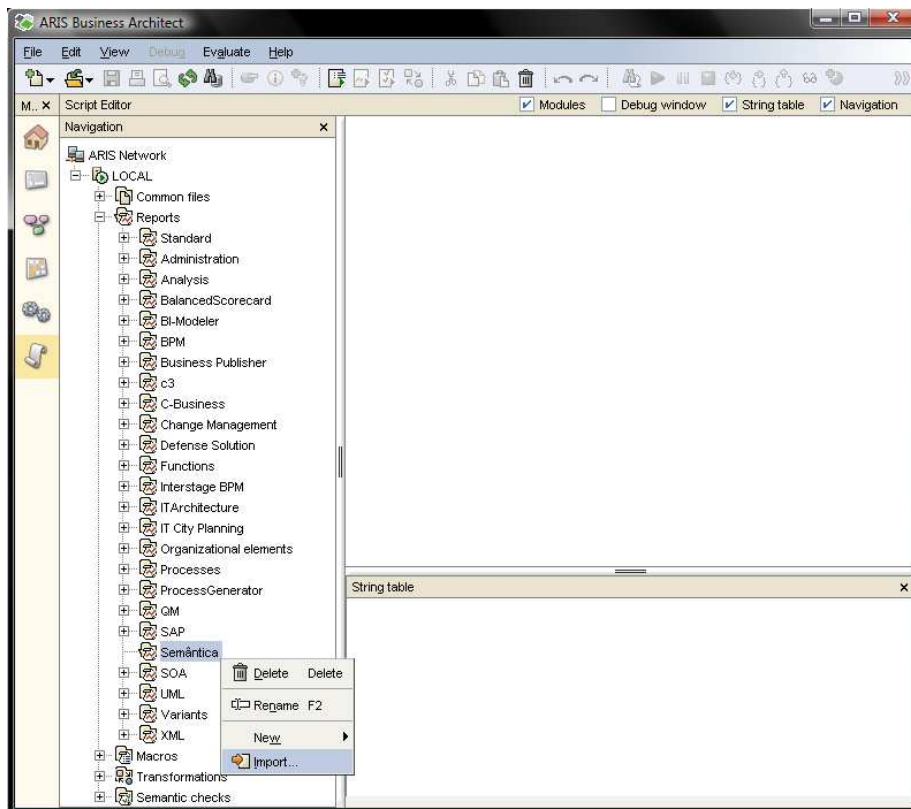


Figura 7 - Importação do script

2.2.2 Conteúdo do script

O script para abertura do Protégé foi desenvolvido em JavaScript. O algoritmo implementado está listado na Figura 8.

```

var objetos = ArisData.getSelectedObjDefs();
var protege = "C:\\Program Files\\Petro-Protege\\Petro-Protege.exe";
var ontologia = "C:\\Perfuracao_20081111\\OWL\\Perfuracao.pprj";
var anotacoes_semanticas = "";
if(objetos.length <= 0){
    var modelos = ArisData.getSelectedModels();
    for(var m in modelos)
        objetos = objetos.concat(modelos[m].ObjDefList());
}
for (var k = 0 ; k < objetos.length ; k++){
    if(objetos[k].Attribute(Constants.AT_USER_ATTR2,
Context.getSelectedLanguage()).GetValue(true) != ""){
        anotacoes_semanticas = anotacoes_semanticas +
objetos[k].Attribute(Constants.AT_USER_ATTR2,
Context.getSelectedLanguage()).GetValue(true) + " "; }
}

if (anotacoes_semanticas.length > 0){
    Context.execute(protege + " " + ontologia + " " + anotacoes_semanticas);}
else{
    Dialogs.MsgBox("A ontologia não será exibida por não conter anotações
semânticas selecionadas.");}

```

Figura 8 - Detalhamento do script

Neste script devem ser feitas algumas observações que podem ocasionar em adaptações no código, quando o script estiver importado em definitivo para o ambiente de produção:

- A variável “protege” deve conter o endereço da instalação do Protégé.
- A variável “ontologia” deve conter o endereço da instalação da ontologia.
- O atributo “AT_USER_ATTR2” deve ser atualizado e considerado o mesmo utilizado conforme a seção 2.1 .

3 Integração de modelos lógicos de dados e ontologias

Para a integração de modelos lógicos descritos no Erwin, foram estudadas as formas de integração disponíveis na ferramenta (seção 3.1). A solução adotada por este projeto está descrita em detalhes na seção 3.2 .

3.1 Formas possíveis de integração do Erwin com Protégé

A ferramenta ERwin foi analisada a fim de encontrar meios de integração com o Protégé. Quatro formas foram avaliadas como provável método de integração, conforme desenho abaixo.

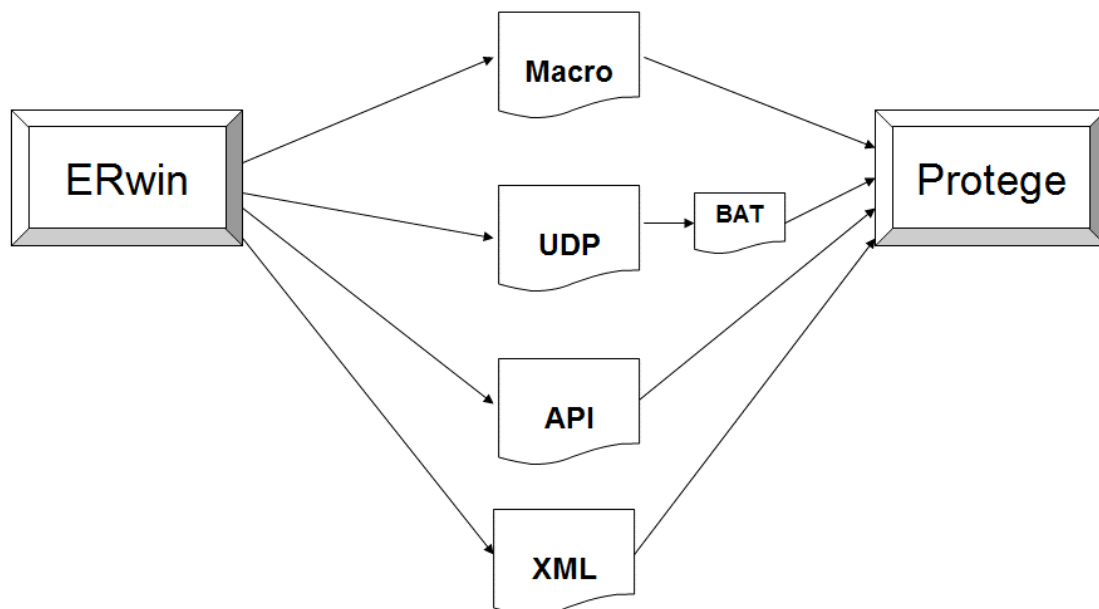


Figura 9 – Prováveis formas de integração ERwin com Protégé

3.1.1 Macro

Para associar o template à tabela basta selecionar no menu a opção Database → Stored Procedures → Table Level ou com o botão direito do mouse, na definição da tabela no digrama, a opção Stored Procedure.

As macros podem ser de cinco tipos:

- Entity Macro - são relacionados às informações das tabelas como nome, relação de atributos, índices, relacionamentos, entre outros.
- Relationship Macro - são relacionadas às informações dos atributos dos relacionamentos entre as tabelas.
- Attribute Macro - são relacionadas às informações dos atributos das tabelas.

- Constraint Macro - são relacionadas às informações de regra, domínio e default do modelo utilizado na definição dos atributos das tabelas.
- Miscellaneous Macro - são macros de variáveis e funções que auxiliam na obtenção e manipulação de informações do modelo como nome do SGBD, nome do database, condições, transformação.

A utilização de macros no ERwin fica limitado ao modelo físico. Essas macros são triggers, stored procedures e scripts que podem ser executados antes ou depois de qualquer alteração feita no modelo de dados. As informações retornadas são sempre referente à tabela associada a macro na stored procedure ou script.

A alternativa de utilizar macros foi desestimulada pelo fato de mexer no modelo físico e o script ser executado quando na criação de uma tabela ou algum tipo de execução desse tipo.

3.1.2 UDP

A User Defined Properties (UDP) é uma propriedade que pode ser de uma entidade ou atributo do modelo do ERWin, e pode ser do tipo texto, numérico, data, lista ou comando. Esse último permite abrir um arquivo. Uma UDP pode ser parametrizada pra cada atributo ou entidade e abrir arquivos diferentes. Porém, as UDPs não possuem a capacidade de passar parâmetros em sua linha de código no momento da abertura do arquivo. A passagem de parâmetros é importante neste projeto pois é necessária para abrir o Protégé customizado. Assim, a alternativa encontrada pra passagem do parâmetro do atributo ou entidade seria abrir um arquivo BATCH e esse ficaria responsável por chamar o arquivo externo e passar o parâmetro por linha de código. O problema dessa alternativa é que, pra cada atributo e entidade, deveria existir um arquivo BATCH diferente. Esse arquivo BATCH deve conter o parâmetro responsável a este atributo ou entidade que o chamou.

O Erwin possui elementos parecidos com UDPs, chamados de add-ins, os quais são possíveis fazer chamada a uma aplicação (exe, bat) ou a um objeto COM. Apesar desse tipo de elemento poder abrir arquivos passando parâmetros, não existe a opção de identificar a entidade ou atributo, pois o componente pertence ao modelo lógico. Assim, a chamada ao arquivo externo é possível, mas não têm como passar parâmetros diferentes para diferentes propriedades.

3.1.3 API

É possível desenvolver APIs para o Erwin utilizando linguagens como o C++, VB.NET dentre outras. Essas API's podem ficar armazenadas em um arquivo compilado de acordo com a linguagem que foi desenvolvida ou, no caso de ser um script, ficar embutida em um arquivo DHTML. Em ambos os casos, a API é armazenada fisicamente fora do aplicativo ERwin.

APIs, por poderem ser programáveis, são altamente flexíveis e poderiam trazer bons resultados para o projeto. Porém, os testes realizados não foram produtivos. O Erwin é distribuído com um manual de referência sobre APIs e alguns exemplos tanto em C++ como em VB.NET. Alguns dos exemplos citados no guia de referência foram testados estão defasados. Alguns dos comandos não são mais utilizados, comandos alternativos foram usados, porém não obtiveram o mesmo efeito do previsto pelo guia de referência de API. Fóruns [CA, 2008b] foram pesquisados em busca de outra alternativa utilizando o C++ para desenvolver a API, porém muito pouco material foi

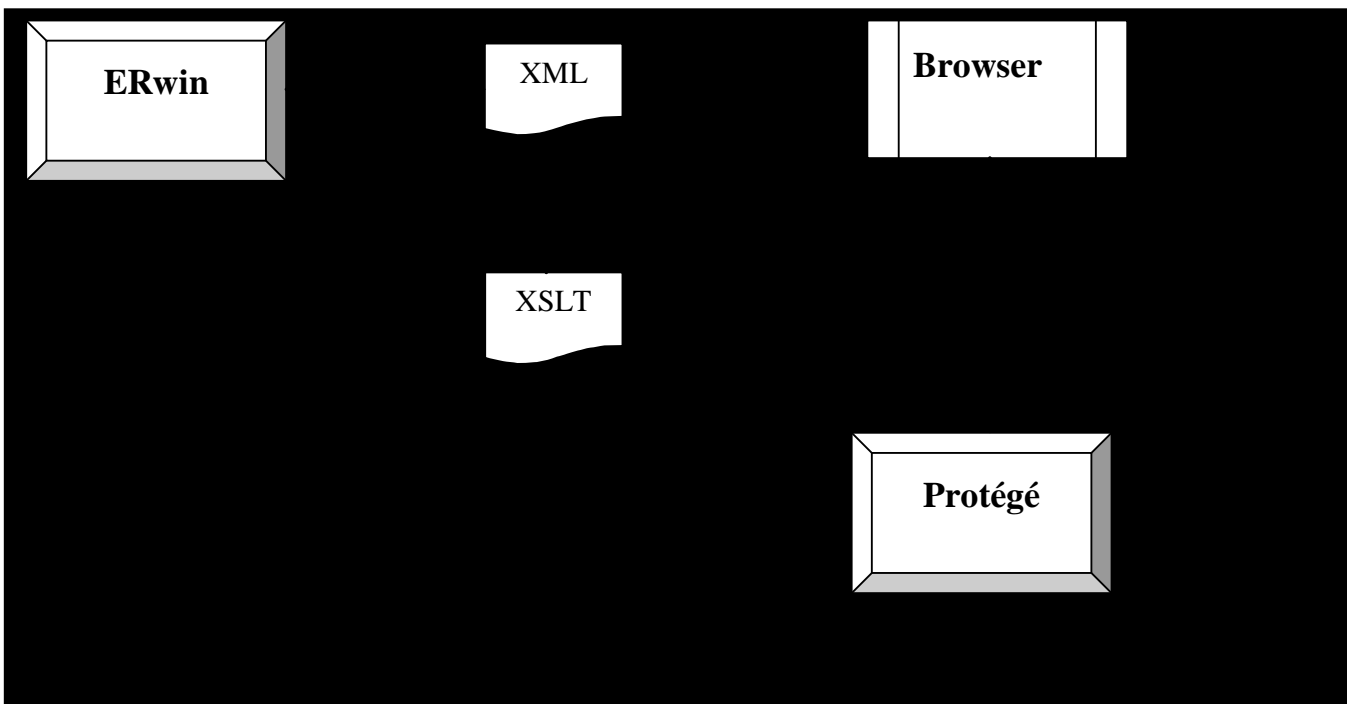
encontrado sobre o mesmo, o que dificultou o desenvolvimento de uma API customizada para o projeto.

3.1.4 XML

O Erwin permite a exportação dos modelos em formato XML. XML é uma metalinguagem atualmente usada, entre outros objetivos, para transportar informação entre sistemas ou ferramentas. A exportação dos modelos para XML não possui restrição e, conforme avaliado no projeto, o arquivo XML gerado possui todas as informações existentes no modelo. A geração de arquivos XML pode ser utilizada como forma de integração externa do Erwin com o Protégé, uma vez que não houve sucesso na criação de uma solução de integração interna à ferramenta.

3.2 Proposta de utilização do Erwin para integração com Protégé

A alternativa mais viável encontrada de integração entre ERwin e Protégé é através da exportação do modelo Erwin em formato XML. O arquivo XML pode ser visualizado como uma página HTML, utilizando a linguagem de transformação XSLT [XSLT, 2008]. Após sua transformação, a página HTML gerada conterá código JavaScript para abrir o Protégé customizado para o projeto, o qual não permite edição da ontologia. As informações semânticas dos elementos do modelo lógico serão anotados utilizando UDPs de tipo texto, intituladas de “Anotação Semântica” e “Namespace da ontologia” referentes às tabelas e aos atributos. A figura abaixo representa esse processo:



por:

- Ler o XML gerado;
- Incluir uma linha de referência a sua folha de estilo XSLT;
- Abrir o XML.

Cada uma dessas etapas é melhor descrita abaixo.

3.2.1 Criação de UDP

Deve ser criado duas UDP's do tipo texto: uma com anotação semântica e outra com o "namespace" da ontologia, ou seja, com o domínio dessa ontologia. Essas duas UDP's devem fazer parte de cada tabela e cada atributo que possua um termo de ontologia correspondente.

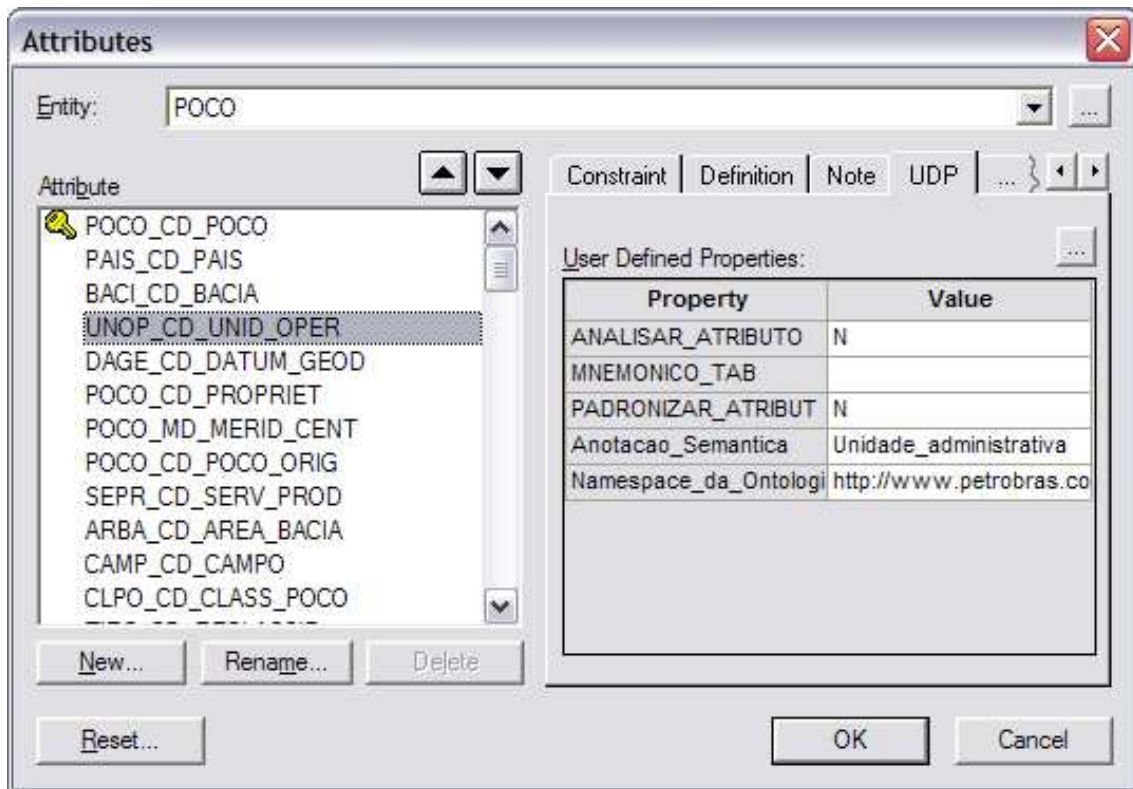


Figura 11 – “Anotação semântica” e “Namespace da ontologia” preenchidos

3.2.2 Exportar modelo lógico para o formato XML

Após o modelo lógico conter as anotações semânticas, este deve ser exportado em formato XML com o nome de ERwin.xml para a mesma pasta onde estará o arquivo que fará sua leitura, conforme mostra a figura abaixo.

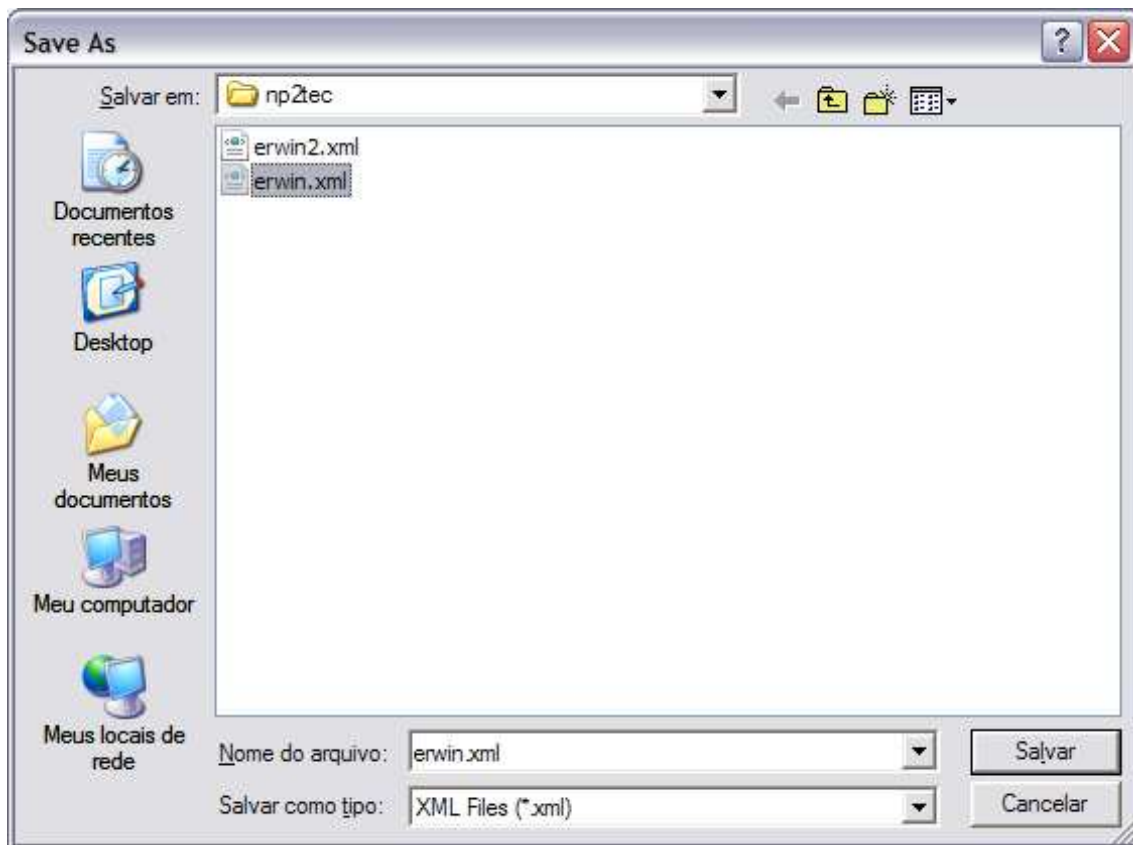


Figura 12 – Salvando como formato XML

3.2.3 Interface que transforma o XML exportado

Um arquivo do tipo XSLT (Extensible Stylesheet Language Transformations), linguagem usada para localização de elementos XML, transformação e formatação, foi desenvolvido para processar o arquivo ERwin.xml e transformá-lo em formato HTML. XSLT é uma linguagem utilizada para transformar vocabulários XML em outros vocabulários XML (em especial, HTML).

Tendo um arquivo XML de origem, o papel do XSLT é transformar esse arquivo em um outro formato, como HTML, XML, PDF ou outro. Assim, tendo um interpretador XML que suporte XSLT, basta escrever um código XSLT que transforme o conteúdo do documento XML em HTML. Observe o diagrama a seguir:



Figura 13 – Diagrama XSLT

Abaixo está representado a tela de exibição do arquivo XML exportado do Erwin e processado pelo documento XSLT produzido no projeto.

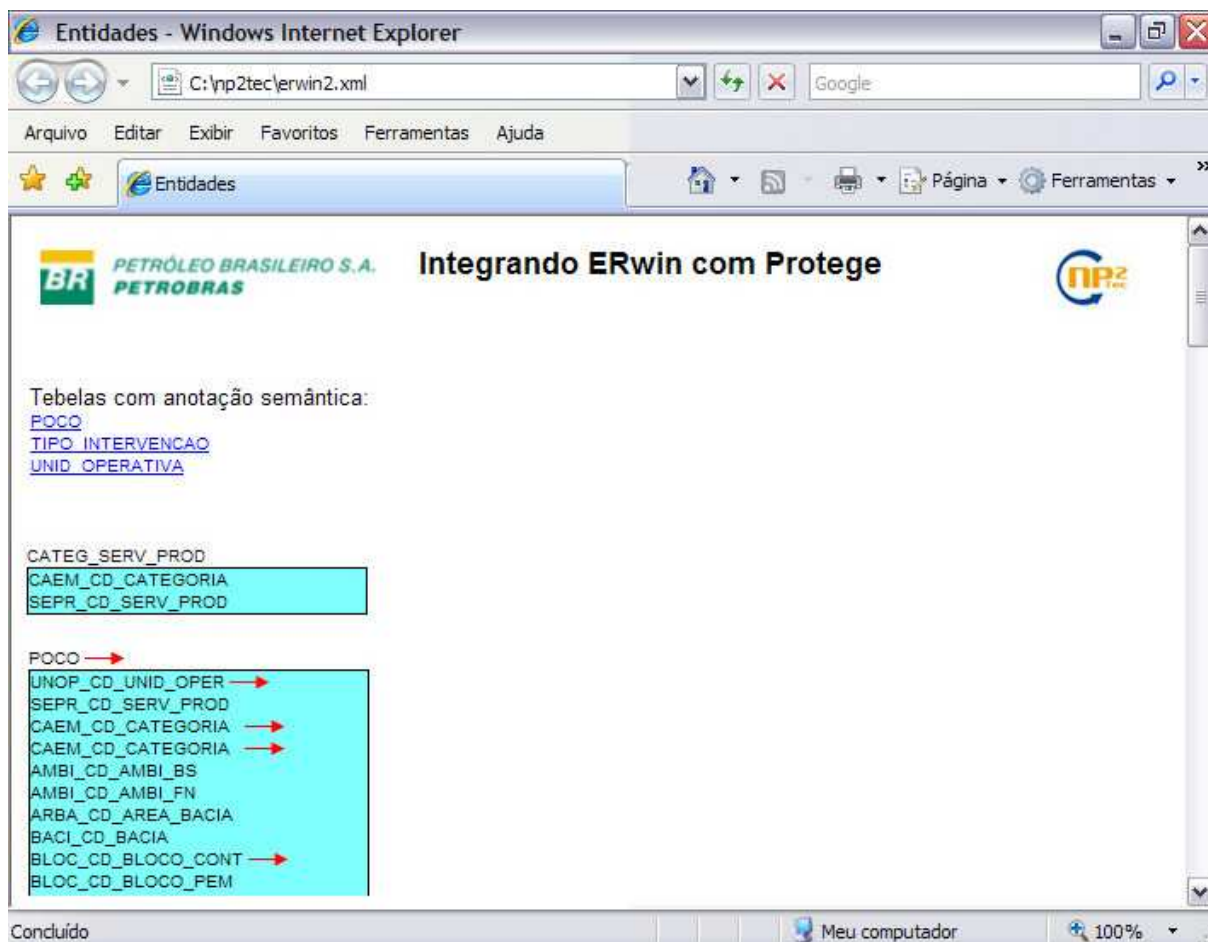


Figura 14 – Interface XML com referência a um XSLT

Existe a necessidade de referenciar o arquivo XSLT no conteúdo do arquivo XML exportado. Foi desenvolvida uma aplicação Java no projeto para inserir essa referência. Essa aplicação lê o XML, verifica sua estrutura e o valida para que possa ser corretamente transformado pelo XSLT, insere a referência para o XSLT e executa o XML num navegador de internet. O arquivo XSLT é responsável por ler o XML e transformá-lo, onde a estrutura de árvore do XML dá lugar à interface em HTML.

O código do arquivo XSLT implementado pode ser verificado na figura abaixo.

```

<?xml version="1.0"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:key name="distinct" match="Name" use="."/>
  <xsl:template match="/">
    <HTML>
      <HEAD>
        <TITLE>Entidades</TITLE>
        <link href="estilo.css" rel="stylesheet"></link>
        <script>
          function abrirOntologia(var1) {
            var command = "C:\\Petro-Prottega_20081117\\Petro-Prottega.exe C:\\owl\\Perfuracao.pprj Boco";
            var shell=new ActiveXObject("WScript.shell");
            shell.run(command);
          }
        </script>
      </HEAD>
      <BODY>
        <img><xsl:attribute name="src">np2tec.jpg</xsl:attribute></img>
        <table width="646" cellspacing="0" cellpadding="0" border="0">
          <tr>
            <td align="center"><h3>Integrando ERwin com Protege</h3></td>
          </tr>
        </table>
        <a onclick="abrirOntologia('teste')" style="cursor: hand;">
          abrir notepad 0
        </a><br/><br/>
        <xsl:for-each select="/ERwin//Entity_Groups/Entity">
          <xsl:sort select="EntityProps/Name" order="ascending"/>
          <table width="100" cellspacing="0" cellpadding="0" border="0">
            <tr>
              <td>
                <a>
                  <xsl:attribute name="href">C:\\Protege_3.3.1\\Protege.exe <xsl:value-of
select="EntityProps/UDP_Instance_Groups/UDP_Instance[5]"/></xsl:attribute>
                  <xsl:value-of select="EntityProps/Name"/>
                </a>
              </td>
            </tr>
            <tr><td>
              <table width="100%" cellspacing="0" cellpadding="1" bgcolor="#000000" border="0" bordercolor="#000000">
                <tr><td>
                  <table width="100%" cellspacing="0" cellpadding="0" border="0" bgcolor="#80ffff">
                    <xsl:for-each select="Attribute_Groups/Attribute">
                      <xsl:sort select="AttributeProps/Name" order="ascending"/>
                      <tr>
                        <td>
                          <a>
                            <xsl:attribute name="href">C:\\Protege_3.3.1\\Protege.exe <xsl:value-of
select="AttributeProps/UDP_Instance_Groups/UDP_Instance[4]"/></xsl:attribute>
                            <xsl:value-of select="@name"/>
                          </a>
                        </td>
                      </tr>
                    </xsl:for-each>
                  </table>
                </td></tr>
              </table>
            </td></tr>
          </table>
          <br/>
        </xsl:for-each>
      </BODY>
    </HTML>
  </xsl:template>
</xsl:stylesheet>

```

Figura 15 – Código do arquivo XSLT

4 Integração de base de dados com ontologias

O banco de dados Oracle é amplamente utilizado no ambiente empresarial. Essa ferramenta incluiu funcionalidades semânticas a partir da versão 10g, com acréscimo de mais funcionalidades na versão seguinte, 11g.

A ferramenta Oracle possui suporte para armazenamento de modelos ontológicos e realização de consultas baseadas nas inferências desses modelos.

A versão Oracle 10g armazena modelos no formato RDF e RDF-S, realizando consultas a partir da estrutura de triplas definida no formato RDF.

A versão Oracle 11g armazena modelos no formato RDF, RDF-S e OWL, realizando consultas a partir das estruturas definidas pelos construtos dessas linguagens, permitindo inferências mais complexas a partir de construtos OWL. A ferramenta também oferece um suporte elementar à execução de consultas dos dados relacionais utilizando o nome de itens presentes nos modelos ontológicos, fazendo uma comparação do nome do item ontológico com o valor de uma coluna de uma tupla na base de dados relacional.

4.1 Detalhamento da ferramenta Oracle 11g

A ferramenta permite a execução de consultas do modelo semântico, retornando informações somente sobre o que está armazenado no modelo semântico da ontologia. Entretanto, possui mecanismos que permitem a utilização de esquemas externos ao esquema semântico, criando restrições às consultas realizadas a partir de comparações do nome de itens da ontologia a itens no esquema externo, usando o predicado de triplas da ontologia para comparação. A inclusão de informações na ontologia ocorre via interface de linhas de comando.

Os testes realizados mostraram que os tipos de consultas permitidas são úteis somente para modelos ontológicos, no caso OWL, e não para consultas aos dados relacionais com base em informações nos modelos ontológicos. As consultas utilizando esquemas externos têm seus resultados limitados, pois somente uma operação de comparação entre strings de caracteres é possível, a partir do valor das tuplas do esquema externo, inexistindo, assim, um mapeamento direto entre a ontologia e os dados relacionais.

Para que uma consulta utilizando os recursos presentes em modelos semânticos, é necessário que os dados do esquema externo estejam presentes como instâncias na ontologia. Caso contrário, não há informação suficiente para realizar as restrições, e a impossibilidade de mapear um item ontológico a um item do esquema externo impede o uso da estrutura dos dados relacionais em consultas com inferências ontológicas.

A ferramenta não possui uma interface gráfica para edição, apresentando uma interface de linha de comando para acrescentar as informações a uma ontologia, dificultando a atividade de manipulação da ontologia, pois não se tem a visualização completa do que está presente na ontologia durante a sua construção. A OWLPRIME, versão OWL suportada pela ferramenta, não suporta todos os construtos OWL e para que as vantagens de inferências OWL sejam aproveitadas, as instâncias das informações devem estar presentes no modelo da ontologia.

4.2 Comunicação com o Protégé

O Protégé possui uma interface com o Oracle através de plugins, permitindo que a criação da ontologia seja facilitada pelo uso de uma interface gráfica e o armazenamento seja realizado no SGBD.

Como mencionado anteriormente, o banco de dados Oracle não possui um mecanismo gráfico de edição de ontologias. Isto dificulta sua construção, pois não é

possível visualizar completamente os itens existentes na ontologia, e, por essa manipulação ser por linha de comando, o acréscimo de novos conceitos é prejudicado.

A ferramenta Protégé possui um plug-in, chamado Oracle RDF [ORACLE, 2008b] e recomendado pela Oracle, que permite a importação e exportação de triplas dos modelos semânticos de ontologias do banco de dados Oracle, com suporte para as versões Oracle 10g e 11g.

Para a importação, a criação de uma nova ontologia é necessária, definindo que o seu tipo é "Oracle RDF Store", ou repositório Oracle de RDF. Esse passo é apresentado na Figura 16.

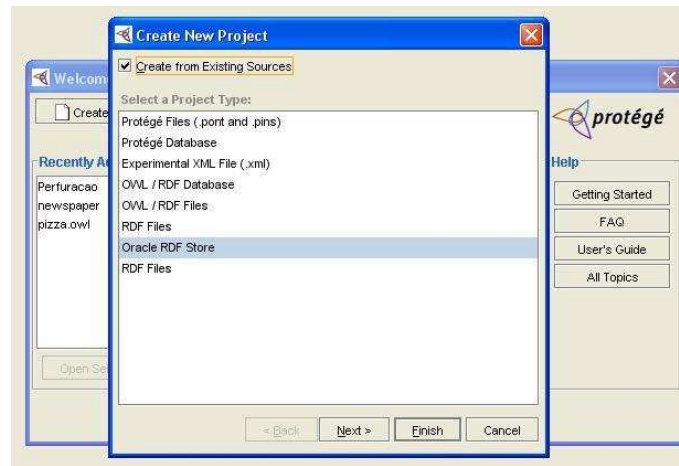


Figura 16 – Passo inicial da importação

Em seguida, as informações sobre a ontologia no banco de dados Oracle devem ser preenchidas. As informações necessárias são o caminho do servidor de banco de dados, incluindo a porta, o identificador do banco de dados, o nome do usuário que tenha privilégios de acesso ao modelo com sua senha, a tabela que contém as informações da ontologia e o namespace da ontologia. Esse passo é apresentado na Figura 17.

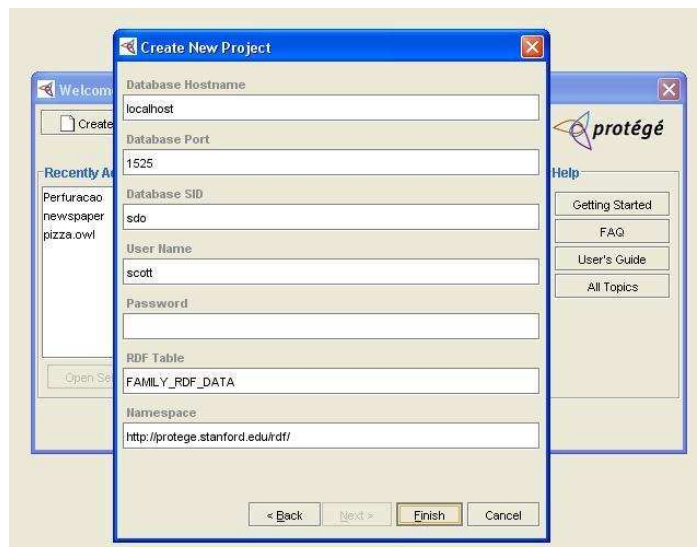


Figura 17 – Definição de informações sobre a ontologia armazenada no Oracle

Para a exportação, é necessário que uma ontologia esteja sendo editada, para que em seguida seja escolhida a opção “Export to Format” e em seguida “Oracle RDF Store” no menu “File. Esse passo é apresentado na Figura 18.

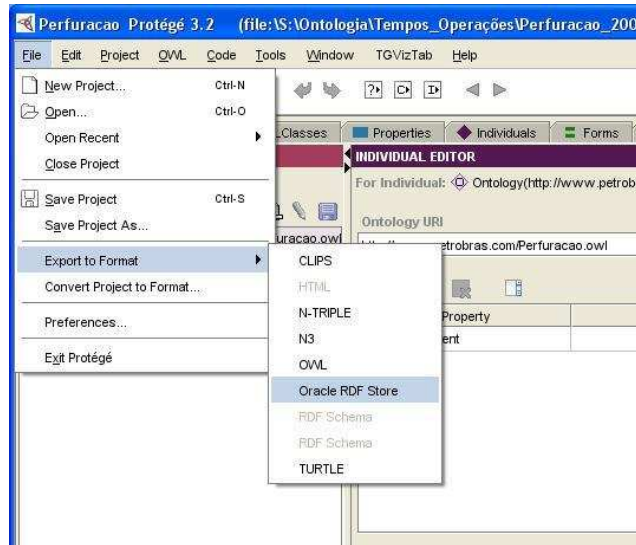


Figura 18 – Passo inicial da exportação

Após esse passo, informações da ontologia no banco de dados Oracle que vai receber as informações devem ser preenchidas. As informações necessárias são o caminho do servidor de banco de dados, incluindo a porta, o identificador do banco de dados, o nome do usuário que tenha privilégios de acesso ao modelo com sua senha e a tabela que receberá as informações da ontologia sendo manipulada no Protégé. Esse passo é apresentado na Figura 19.



Figura 19 – Definição de importações para exportação da ontologia

Esses são os passos para a exportação e importação de ontologias do banco de dados Oracle pelo Protégé. Entretanto, esse processo possui limitações: somente suporta formato RDF, impedindo que um modelo OWL não seja armazenado, e não atualiza os dados durante a exportação, ou seja, caso seja necessário atualizar a

ontologia no repositório, os dados são incluídos no fim da tabela, ficando repetidos. Esse conjunto de limitações torna o uso do plug-in na sua atual versão inviável.

5 Considerações sobre as soluções propostas de integração

A solução proposta no projeto de Integração de Modelos com Ontologias teve como finalidade mostrar, em um curto período de tempo, a viabilidade de integração entre as ferramentas ARIS, Erwin e Oracle com a ferramenta Protégé. Esta seção discute as limitações dos protótipos implementados e apresenta soluções para tais limitações. Ainda, a seção discute os pontos de alerta para implantação do protótipo.

5.1 Falta de um repositório de ontologias

Uma primeira limitação a ser destacada é a falta de um repositório para armazenamento das ontologias. O protótipo desenvolvido resolve o problema de localização da ontologia fixando um endereço padrão para a ontologia, descrito na seção 2.2.2. O cenário ideal para o armazenamento e localização de ontologias seria o uso de um repositório com a finalidade de armazenar apropriadamente ontologias. Tal repositório deve permitir a identificação de ontologias não pelo endereço de arquivo, como é o caso da proposta atual, mas sim através de um identificador único. Um possível identificador para a ontologia poderia ser o espaço de nomes (namespace) que a ontologia descreve, caso todas as ontologias criadas na corporação descrevam um, e somente um, espaço de nomes. O atributo “Namespace da ontologia” foi definida visando essa solução. A figura seguinte ilustra o uso e as funções do repositório de ontologias.

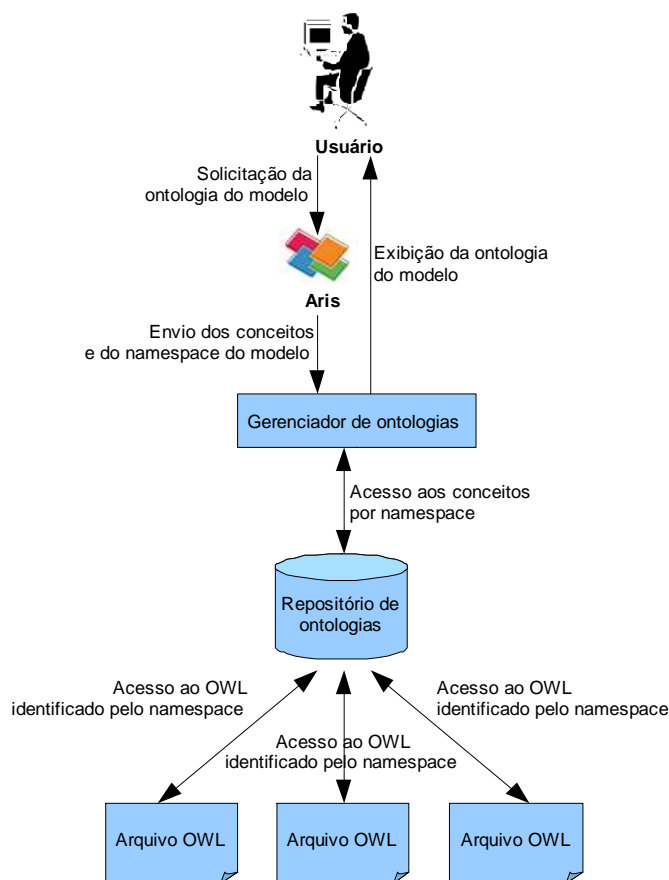


Figura 20 - Acesso a ontologias através de um repositório

5.2 Definições de infra-estrutura necessárias à execução da solução proposta

O script da seção 2.2.2 possui variáveis que armazenam o endereço da instalação do Protégé customizado e o endereço da ontologia de domínio que possuem valor constante. No script as variáveis “protege” e “ontologia” precisam ser adaptadas para o endereço em que serão feitas as instalações nos computadores. É preciso ter uma atenção a cada vez que forem feitas as instalações, para identificar se os diretórios de instalação estão de acordo com os esperados pelo script.

Por fim, a solução de integração desenvolvida faz uso de dois atributos. Para a solução criada para o ARIS, estes atributos possuem, por padrão, nomes no formato AT_USER_ATTRx, sendo x um inteiro qualquer. Contudo, ainda que a ferramenta ARIS permita alterar o nome de tais atributos, fisicamente o nome do atributo continua inalterado, ou seja, a alteração do nome do atributo funciona somente como uma máscara de apresentação. Assim, uma vez implantado o script em um servidor ARIS, é necessário verificar se o “API name” utilizado para a propriedade “Anotação semântica” condiz com o “API name” da variável definida no script. Para a ferramenta Erwin, não há problema semelhante.

5.3 Limitação de licenças para o ARIS

A solução proposta exige algumas alterações no ARIS para facilitar o modelador de processos anotar e compreender a semântica dos elementos anotados. Sabe-se, contudo, que a disponibilidade de licenças para manipulação da ferramenta ARIS é limitada e que é prática comum na Petrobras publicar a modelagem de processos em páginas HTML, geradas pela própria ferramenta ARIS, na Intranet da organização.

Para disponibilizar a integração com o Protégé para usuários sem licença de acesso à modelagem de processos no ARIS, pode-se utilizar o módulo web publisher para exportar a modelagem de processos em formato HTML, porém exibindo os valores dos atributos criados “Namespace da ontologia” e “Anotação semântica”, conforme mostrado na figura seguinte.

The screenshot shows the E&P Modelagem de Processos interface. On the left, there is a sidebar with an 'Atributos' (Attributes) table. The main area displays a BPMN diagram titled 'Editar ordem de intervenção de manutenção' (Edit maintenance intervention order). The diagram includes a central task 'Editar ordem de intervenção de manutenção' connected to other tasks like 'Intervenção', 'Limpeza', and 'Alto do Rodrigues'. External entities like 'UN-RNCE/ATP-AR', 'Coordenador da manutenção', and 'SAP-PM' are also shown.

Atributos	
Identificação	Intervencao
Semântica	Intervencao
Criador	system
Description	Termo em inglês: Intervention. Qualquer serviço de manutenção, operação ou inspeção que tenha influência nas condições operacionais de equipamentos e sistemas.
Identifier	Intervencao
Last change	09/12/2008 13:30:20
Last user	system
Name	Intervencao
Responsabilidade da ontologia	http://www.petrobras.com/Penuracao.owl
Time of generation	27/10/2008 10:02:25
Type	Termo técnico

Figura 21 – Exportação HTML exibindo o valor dos atributos criados

Uma vez exportada a base de processos em formato HTML, é possível, por exemplo, criar um plug-in para o navegador Mozilla Firefox que leia os valores dos atributos criados na anotação e abra o Protégé com o conceito correspondente na ontologia. Essa solução não exige alteração do processo de exportação em HTML, porém exige a instalação do plug-in criado em todas as máquinas dos usuários que forem interagir com essa exportação HTML. Uma outra solução é a alteração do template HTML utilizado pelo ARIS para exportar a base de processos. Essa alteração do template HTML adicionaria o código JavaScript para chamada do Protégé via navegador e acrescentaria na interface HTML um botão ou link para acionar o código JavaScript. Contudo, é necessário uma análise mais aprofundada do módulo web publisher do ARIS para avaliar as limitações do módulo.

Uma vez que a solução adotada no Erwin já prevê a exportação do modelo em XML e visualização em HTML, não existem problemas de limitação de licenças.

5.4 Limitações da comunicação com o Oracle

A interface entre o Protégé e Oracle via plug-in "Oracle RDF Plugin for Protégé" não se mostrou viável porque está limitada a linguagem RDF, impossibilitando a sua utilização em conjunto de ontologias em OWL. Além disso, o plugin não oferece funcionalidades básicas como a atualização de ontologias no repositório. O banco de dados Oracle 10g não tem suporte OWL, limitando o seu uso com ontologias com maior semântica. O banco de dados Oracle 11g possui suporte parcial à linguagem OWL. Tais limitações fazem com que o armazenamento de ontologias OWL do Protégé no Oracle sofram perdas consideráveis de representação.

6 Conclusões

Este documento relatou o resultado do estudo da integração de modelos com ontologias, apresentando soluções de comunicação entre as ferramentas Aris, Erwin e Oracle com a ferramenta Protégé. O estudo mostra a viabilidade de integração dessas ferramentas através do desenvolvimento de scripts, avaliações de plugins e utilização de uma versão da ferramenta Protégé que possua somente funcionalidades de visualização.

As soluções para as ferramentas ARIS e Erwin permitem visualizar todos os conceitos de uma ontologia de domínio que estão associados a um determinado elemento do modelo de processos do negócio ou do modelo lógico de dados. Essa possibilidade de associação permite uma melhor interpretação dos modelos com a ajuda do conhecimento do domínio descrito pelos conceitos através da ontologia. Desta forma, consegue-se agregar uma melhor qualidade e análise dos processos e das modelagens lógicas e físicas de dados, além de uma melhor integração desses modelos com o domínio de conhecimento da área e os seus termos definidos consensualmente.

Referências Bibliográficas

- CA, **ERwin Data Modeler**, 2008. Disponível em <http://www.ca.com/us/products/product.aspx?id=260>. Último acesso em 26 nov. 2008a.
- CA, **User Community Online Forums**. Disponível em <http://caforums.ca.com/ca/>. Último acesso em 09 dez. 2008b.
- CARDOSO, J., **The Semantic Web Vision: Where are we?**. In: IEEE Intelligent Systems, vol 22, n 5, pp 84-88. 2007.
- IDS-SCHEER, **ARIS Plataforma**. Disponível em <http://www.ids-scheer.com/pt/ARIS/127331.html>. Último acesso em 25 nov. 2008.
- PROTÉGÉ, **The Protégé Ontology Editor and Knowledge Acquisition System**. Disponível em <http://protege.stanford.edu>. Último acesso em 25 nov. 2008.
- ORACLE, **Semantic Technologies Center**. Disponível em http://www.oracle.com/technology/tech/semantic_technologies/index.html. Último acesso em 15 dez. 2008a.
- ORACLE, **Oracle RDF Schema Storage for Protege**. Disponível em http://www.oracle.com/technology/tech/semantic_technologies/sample_code/index.html. Último acesso em 13 dez. 2008b.
- OREN, E., MÖLLER, K., SCERRI, S., HANDSCHUH, S., SINTEK, M. (2006). **What are Semantic Annotations?**. Relatório Técnico. Disponível em <http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf>. Último acesso em 08 dez. 2008.
- XSLT, **Extensible Stylesheet Language Transformations**. Disponível em <http://www.w3schools.com/xsl/>. Último acesso em 26 nov. 2008.