

ROSAI - Uma abordagem baseada em lógica para representar e consultar objetos de aprendizado

Francisco H. T. Costa¹, Fábio A. M. Porto²

Instituto Militar de Engenharia – IME-RJ¹
Departamento de Engenharia de Sistemas – Rio de Janeiro – Brazil

Ecole Polytechnique Fédérale de Lausanne – EPFL²
Faculté Informatique et Communications – LBD Lausanne - Suisse

emails: handrick@de9.ime.eb.br, fabio.porto@epfl.ch

Resumo

O ROSA é um sistema para recuperação de objetos de aprendizado baseado na descrição semântica de seus conteúdos. Este trabalho tem por objetivo estender o modelo de dados ROSA, a partir de sua representação como uma base de conhecimento, na quais assertivas são definidas como fatos e restrições e regras são expressas através de fórmulas lógicas. Consultas são expressões lógicas com variáveis livres. Sua avaliação sobre a base de conhecimento permite a inferência de novos fatos, segundo a premissa do mundo aberto. A base de conhecimento é implementada em uma máquina Prolog que tratará fatos e regras, correspondendo ao domínio de aplicações suportadas pelo ROSA.

1. Introdução

A Web Semântica, conforme proposta por Berners-Lee [01], sugere o desenvolvimento de tecnologia para suporte à descrição e processamento semântico de recursos disponíveis na Web, tais como (XML, RDF, DAML-OIL, OWL, etc). Um aspecto relevante deste elenco de tecnologias está na possibilidade de representação e processamento de relacionamentos rotulados entre recursos. Tais relacionamentos estabelecem associações semânticas entre conceitos ou recursos, segundo diferentes níveis de qualificação, dependendo da tecnologia empregada (RDF, DAML-OIL, etc).

Diversas aplicações se mostram beneficiárias de um modelo de dados no qual a representação semântica de suas associações seja computável. Essas associações são relevantes tanto na descrição dos objetos do domínio quanto no suporte à consultas de usuários. Ferramentas de gerência de documentos compartilhados (ex. arquivos de músicas) constituem um exemplo dessas aplicações.

ROSA[2] é um sistema concebido segundo este princípio, no qual objetos de aprendizagem são contextualizados por redes semânticas que associam: Cursos, Disciplinas e Tópicos. ROSA segue a abordagem de banco de dados para representação de seu modelo e consultas são expressas diretamente sobre o modelo lógico.

Este trabalho tem por objetivo estender o sistema ROSA a partir da representação de seu modelo através de uma base de conhecimento. Para tal, o modelo de dados foi representado como

fatos na base de conhecimento e enriquecido com a definição de regras. Consultas, como as expressas em ROSA atual, foram traduzidas para fórmulas lógicas com variáveis livres. Em sua avaliação, consultas fazem uso de inferência sobre os fatos e regras permitindo inferir fatos não definidos na base (premissa do mundo-aberto), bem como tratar herança de conceitos e propriedades de relacionamentos, de forma automática.

O restante desse trabalho encontra-se estruturado da seguinte forma. Na Seção 2 é feita uma breve descrição do sistema ROSA, destacando-se sua estrutura atual e identificando pontos interessantes para sua extensão. Em seguida, na Seção 3, apresenta-se a representação em lógica para o modelo de dados ROSA. A Seção 4 discute a arquitetura proposta para o ROSAI e na Seção 5 são descritos alguns detalhes de implementação. A Seção 6 destaca iniciativa similar a do ROSAI e, finalmente na Seção 7 apresentam-se as considerações finais.

2. O Sistema ROSA

O sistema ROSA visa auxiliar profissionais da área educacional, ajudando-lhes a descobrir conteúdos didáticos armazenados no sistema, e desta forma, fornecendo subsídios para preparação de suas aulas ou conteúdos instrucionais. Para isso o sistema armazena Objetos de Aprendizado, que representam de fato os conteúdos instrucionais, acrescidos de um conjunto de características e propriedades, também denominadas de metadados, além de um conjunto de associações (ou predicados) que expressam os relacionamentos entre os objetos. O leitor interessado pode encontrar detalhes sobre o modelo de dados em [10]. A figura abaixo mostra a arquitetura atual do sistema ROSA.

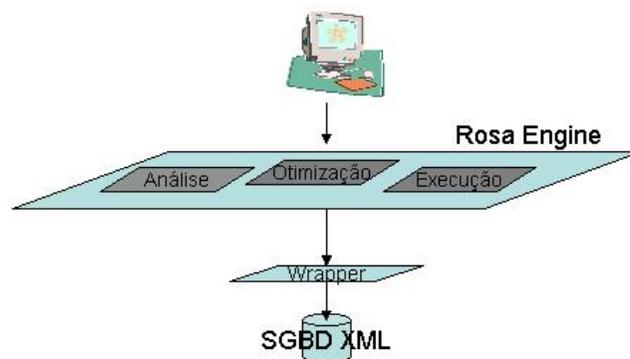


Figura 1: Arquitetura do Sistema ROSA

O processamento de consultas no modelo ROSA compreende as etapas tradicionais acrescidas de um adaptador para conversão entre os modelos de dados XML, de armazenamento, e ROSA. O sistema recebe uma consulta, expressa na linguagem ROSAQL, realiza as etapas iniciais de normalização e produz uma estrutura intermediária. Em seguida, invoca os serviços de geração de plano, que produz uma expressão algébrica utilizando-se dos operadores da álgebra e algumas heurísticas de otimização. O plano assim produzido é submetido à máquina de execução de consultas que acessa fontes de dados e avalia as operações contidas no plano. Na versão atual, apenas a máquina de execução, o acesso a fonte de dados e a interface encontram-se disponíveis.

De modo a permitir o tratamento de inferência, este trabalho pretende substituir a máquina de execução atual por uma máquina de inferência. Esta última avaliará regras armazenadas em uma base de conhecimento, como mostra a figura 2 seguir:

| Linguagem | Tradução para o modelo | Processamento da Consulta |
|-----------|-------------------------------------|--|
| Lógica | Regras + Fatos + Máq. de Inferência | Otimização de consultas mais complexas + Avaliação |
| ROSAQL | Álgebra + ROSAQL | Otimização de consultas complexas + Avaliação |

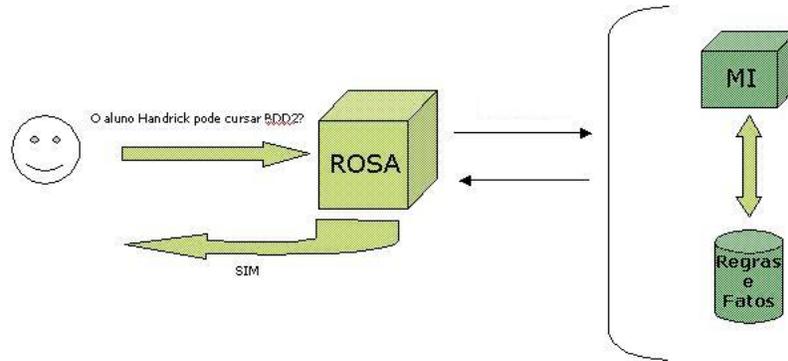


Figura 2: Execução de consultas no modelo ROSA

3. Modelo de Dados

O modelo de dados utilizado no trabalho tem como base Objetos de Aprendizado (Learning Objects – LO) [04]. Consideramos objetos de aprendizagem em dois níveis, lógico e físico. O primeiro inclui disciplinas, cursos e tópicos. O segundo corresponde às mídias digitais propriamente ditas. Objetos de aprendizado são descritos segundo um conjunto de metadados (LOM – IEEE) de cunho educacional. De modo a contextualizá-los, LOs são associados entre si por relacionamentos (ou predicados). Este modelo pode ser visto como uma extensão do modelo RDF [05], onde também é previsto relacionamentos n-ários, cardinalidades, atributos e diferentes tipos de coleções. A idéia básica aqui proposta é representar LOs, e seus relacionamentos, através de uma linguagem lógica. A representação gráfica de um LO e de suas associações é ilustrada abaixo:

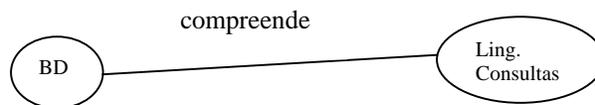
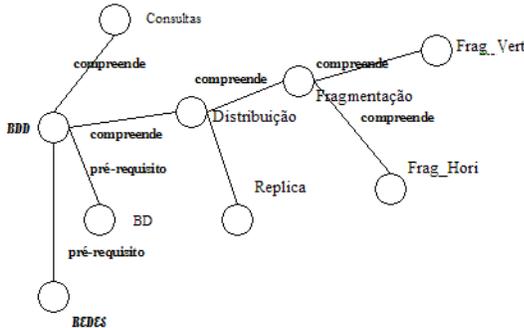


Figura 3: Ligação entre LO(s)

O significado semântico para a associação entre LOs da Figura 3, determina que Disciplina *BD* compreende o Tópico, *linguagens de consulta*. Em uma linguagem lógica, associações são expressas como fatos. O exemplo da Figura 3 pode ser expresso como abaixo:

`'compreende(bd,Linguagem_consulta)'`.

Vista essa pequena representação, fica claro que é possível criar a representação lógica completa de um sistema que segue esse modelo. Além da descrição de fatos, linguagens lógicas permitem expressar consultas sobre a base de fatos e regras. Para isso, variáveis assumem valores que atendem aos predicados definidos na consulta. A seguir é apresentado um pequeno mapa conceitual do modelo ROSA juntamente com sua representação em uma linguagem lógica, no caso, Prolog.



Programa Prolog – Fatos (axiomas)

```
prerequisito(bd,bdd).
prerequisito(redes,bdd).

compreende(bdd,distribuicao).
compreende(distribuicao,fragmentacao).
compreende(fragmentacao,frag_h).
compreende(fragmentacao,frag_v).
compreende(distribuicao,replica).
compreende(bdd,consultas).
```

O modelo lógico também prevê representações de classes como LOs e relacionamentos, bem como suas instâncias de classes com seus devidos atributos. Mais uma vez, as representações são feitas através de axiomas, como pode ser visto nos quadros a seguir:

| Classe LO | Instâncias da Classe LO | Definição dos atributos |
|-----------|--|---|
| lo(X). | lo(bd). lo(redes). lo(computacao). | bd(difícil,disciplina,pt-br). redes(difícil,disciplina,pt-br). computação(difícil,curso,pt-br). |

Quadro 1: Representação dos LOs

| Classe relacionamento | Instância de relacionamento | Definição de relacionamento |
|----------------------------|-------------------------------------|-----------------------------|
| Relacionamento(X,[lo,lo]). | relacionamento(compreende,[lo,lo]). | compreende(computacao,bd). |

Quadro 2: Representação dos relacionamentos

Formular algumas consultas nesse modelo é simples, eficiente e pode trazer um enorme poder de expressividade. Apresentaremos aqui alguns exemplos que demonstram isso. Primeiramente, desejamos realizar uma consulta que obtenha todos os pré-requisitos do Tópico: *Banco de Dados Distribuídos*. Essa consulta é bastante simples e requer a análise dos fatos descritos na base de conhecimento, tal como expresso no código a seguir.

```
prerequisito(bdd, X). → X = (bd; redes).
```

Um exemplo um pouco mais avançado, que necessita realizar a análise de regras é apresentado a seguir. Digamos que se deseje saber quais os assuntos que compreendem o Tópico *Banco de Dados Distribuídos*. Para resolver essa consulta, se faz necessário analisar uma determinada regra fazendo uso de recursão, de modo a explicitar o fato de que um assunto compreende um outro, se ele compreende diretamente aquele, ou quando existe uma cadeia de assuntos entre eles. Na linguagem Prolog, essa regra pode ser descrita da seguinte maneira:

```
com(X, Y) :-compreende(X, Y); (compreende(X, Z), com(Z, Y)).
```

Assim, a consulta proposta, formulada em Prolog, seria descrita pela expressão $com(X,Y)$, tendo como primeiro parâmetro a disciplina da qual se deseja obter os assuntos por ela compreendidos. O exemplo ficaria da seguinte forma: $com(bdd,Y)$. O resultado de tal consulta pode ser visto a seguir:

```
Y=(distribuicao; consultas; fragmentacao; replica; frag_hori; frag_vert).
```

Adicionalmente, esta proposta estende o modelo ROSA com o tratamento de herança de conceitos e relacionamentos. Como exemplo, considere os relacionamentos "agrega" e "contém". Ambos se comportam como o tipo de relacionamento "compreende". Assim o usuário poderá expressar relacionamentos utilizando o termo que melhor lhe convier, bastando para isso associá-lo ao tipo de relacionamento apropriado. Considere o exemplo a seguir:

```
transitivo(Y):-ehconforme(X,Y),transitivo(X).
transitivo(compreende).
ehconforme(compreende,agrega).
ehconforme(compreende,contem).
```

Desta forma, ao avaliar predicados *agrega* e *contem*, o sistema é capaz de associar seu comportamento ao de relacionamento *compreende*, inferindo um comportamento transitivo aos primeiros.

4. Arquitetura do ROSAI (ROSA – Inference)

ROSAI está estruturado em três camadas, a de apresentação, a de inferência e a de persistência dos LOs, ver figura 4.

A camada de interface permite ao usuário submeter consultas ao sistema sem que precise conhecer detalhes de linguagem lógica. Através de uma interface baseada em formulários, usuários podem determinar predicados sobre os metadados de LOs e exprimir navegações.

A camada de persistência armazena os LOs e seus relacionamentos em um documento XML. Através de um tradutor, o sistema converte o modelo expresso em XML, em fatos na base de conhecimento. Um pequeno exemplo do resultado final do documento lógico criado a partir desse documento XML pode ser visto logo a seguir.

| | |
|---|---|
| <pre><los> <lom> <identifier>001</identifier> <general> <title>computacao</title> <language>pt-br</language> <keyword>curso_mestrado</keyword> <difficulty>dificil</difficulty> <aggregationlevel>curso</aggregationlevel> </general> <description>curso_de_mestrado_presente_no_IME</description> <associations> <predicate> <name>compreende</name> <objects> <id>002</id> </objects> </predicate> </associations> </lom> <lom> <identifier>002</identifier> <general> <title>banco_dados</title> <language>pt-br</language> <keyword>tecnologia_armazenamento</keyword> <difficulty>dificil</difficulty></pre> | <pre>lo(computacao). Computacao(001,dificil,curso_mestrado, Curso,pt-br,curso_de_mestrado_presente_no_IME). Compreende(computacao,banco_dados). lo(banco_dados). Banco_dados(002,dificil,tecnologia_armazenamento,disciplin a,pt-br,disciplina_de_mestrado_no_IME). Relacionamento(compreende,[lo,lo]).</pre> |
|---|---|

```

<aggregationlevel>disciplina</aggregationlevel>
<description>disciplina_de_mestrado_no_IME</description>
</general>
</lom>
</los>

```

DOCUMENTO XML

DOCUMENTO LÓGICO

O documento final formará a base de conhecimento e irá trabalhar em conjunto com outro documento lógico, que irá conter todas as regras do ROSAI. O modelo ROSAI estende o modelo ROSA com regras que serão capazes de inferir novo conhecimento.

Como dito anteriormente, consultas no ROSAI são expressas a partir de uma interface de alto nível. Os predicados definidos pelos usuários são reescritos como fórmulas lógicas (conforme Seção 3) e tratados pelo módulo *Gerenciador de Consultas*. Este aciona os módulos de base de conhecimento e de regras para atender a uma consulta.

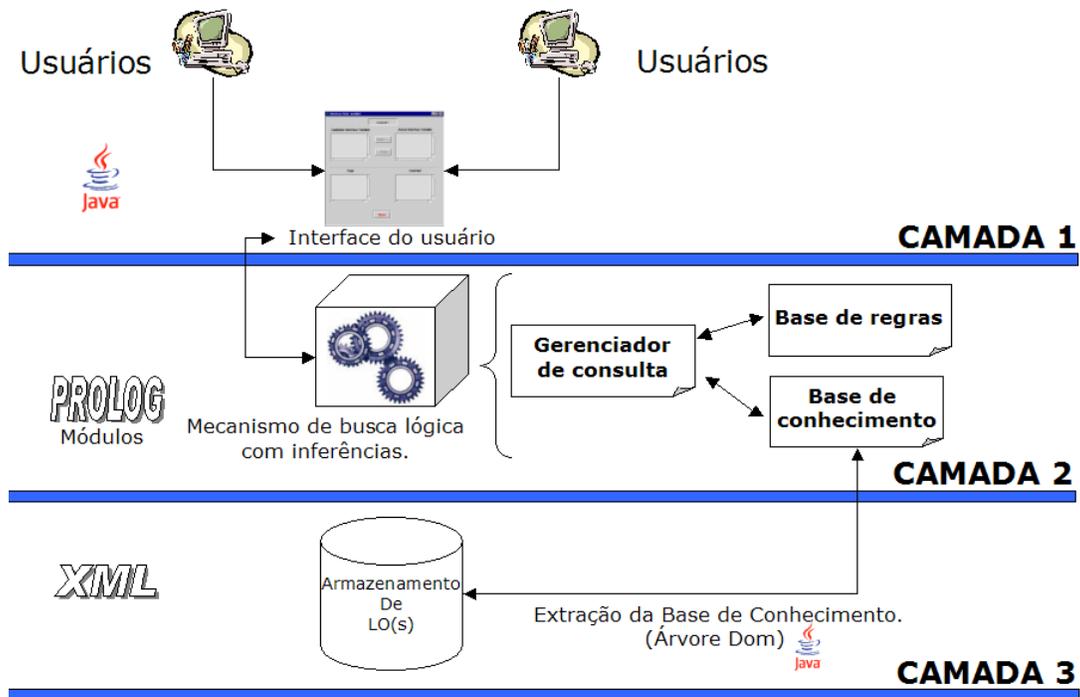


Figura 4: Arquitetura do ROSAI

Quaisquer regras que se façam necessárias definir, tais como: de simetria, transitividade, reflexividade, ou outras que sejam necessárias deverão ser implementadas na base de regras, e quanto melhor a sua elaboração, maior poder de inferência será dada ao sistema.

Um exemplo de inferência no ROSAI pode ser acompanhado, em uma consulta que se deseja conhecer a lista de pré-requisito de uma dada disciplina, como, por exemplo, *Banco de*

A linguagem utilizada na implementação de ROSAI oferece o mesmo poder de expressividade da linguagem lógica Prolog. Em nossos experimentos, todas as consultas propostas no projeto ROSA foram implementadas e, além disso, adicionaram-se consultas que exploram o poder de inferência.

A arquitetura proposta já se encontra com seu principal protótipo implementado permitindo que uma análise de vantagens e desvantagens possa ser realizada. Nesse momento, as consultas propostas em [2] foram implementadas utilizando o modelo e a linguagem Prolog, mostrando o potencial de expressividade da plataforma.

Entre os aspectos que restam a ser verificados destaca-se o comportamento do sistema frente a uma grande base de conhecimento.

Finalmente, como continuação de nosso trabalho, pretendemos avaliar a integração do ambiente lógico com o modelo ROSA, baseado em banco de dados. Tal integração talvez ofereça o melhor dos dois paradigmas.

Referências

- [01] Berners-Lee, T., *The Semantic Web - XML2000*.
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>
- [02] Porto, F., Moura, A., Fernandez, A., Fernandes, A., Coutinho F., Campos, G., Coutinho, L. *ROSA: A Data Model and Query Language for e-Learning Objects*. I PGL DB Research Conference.
- [03] OWL – Ontology Web Language.
Specification. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- [04] LOMWG – IEEE LOM Working Group. [Http://ltsc.ieee.org/wg12/](http://ltsc.ieee.org/wg12/)
- [05] *Resource Description Framework (RDF) Model and Syntax*
Endereço. 1999. <http://www.w3.org/TR/PR-rdf-syntax/>, 1999.
Último acesso Oct. 2001.
- [06] B. Ludascher, Guizhen Yang, and Michael Kifer. FLORA: The secret of object-orienté logic programming. Technical report, SUNY at Stony Brook, 1999.
- [07] Stefan Decker, and Michael Sintek. A linguagem TRIPLE –
Endereço. <http://triple.semanticweb.org/>
Último acesso Fev. 2004
- [08] SWI-Prolog
Endereço. <http://www.swi-prolog.org/>
Último acesso jan. 2004
- [09] Projeto ROSA (Repository of Objects with Semantic Access for e-learning)
Endereço: (www.ipanema.ime.br/~Rosa).
Último acesso maio. 2004
- [10] Query Processing in ROSA Data Model.
Fábio J. C. Silva, Fábio A. M. Porto; SBBD2004. Out. 2004, Brasília. I PGL DB Research Conference, PUC-Rio, Abril, 2003
- [11] ROSA: A Repository of Objects with Semantic Access for e-learning
Fábio Porto, Ana Maria Moura, Fábio José Coutinho da Silva,
IDEAS2004, Coimbra, Julho 7-9,2004
- [12] Project EDUTELLA - RDF-based Metadata Infrastructure for P2P Applications
Endereço: (<http://edutella.jxta.org>)
Último acesso jun. 2004
- [13] Java language
Endereço: (<http://java.sun.com>)
- [14] JPL project
Endereço: (<http://sourceforge.net/projects/jpl/>)