

Um Extrator de Topic Maps a partir de Recursos Heterogêneos de Informação

Giovani Rubert Librelotto
José Carlos Ramalho
Pedro Rangel Henriques

{grl,jcr,prh}@di.uminho.pt

Universidade do Minho - Braga - Portugal

Motivação

- Suponha que exista um conjunto de recursos de informação;
- Deste conjunto, pretende-se extrair uma ontologia:
 - Editando manualmente, o trabalho se torna consumidor de tempo e recursos humanos;
 - Criando um extrator automático, o mesmo estaria apto para extrair a ontologia apenas de um recurso;
 - Então, porque não definir uma especificação para a geração automática de extratores de ontologias?



Index

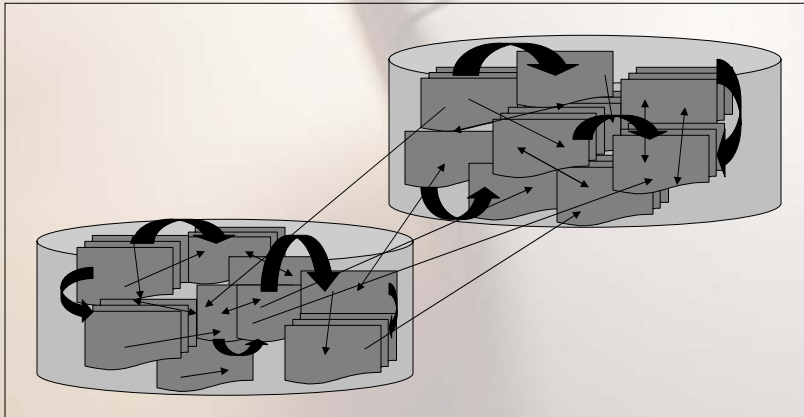
- ❖ **Semantic Web e Ontologia**
- ❖ ***Topic Maps e XTM (XML Topic Maps)***
- ❖ **Metamorphosis**
 1. **TM-Builder – XSTM**
 2. **Oveia – XS4TM**
 3. **Ulisses**

Problemas com motores de busca atuais

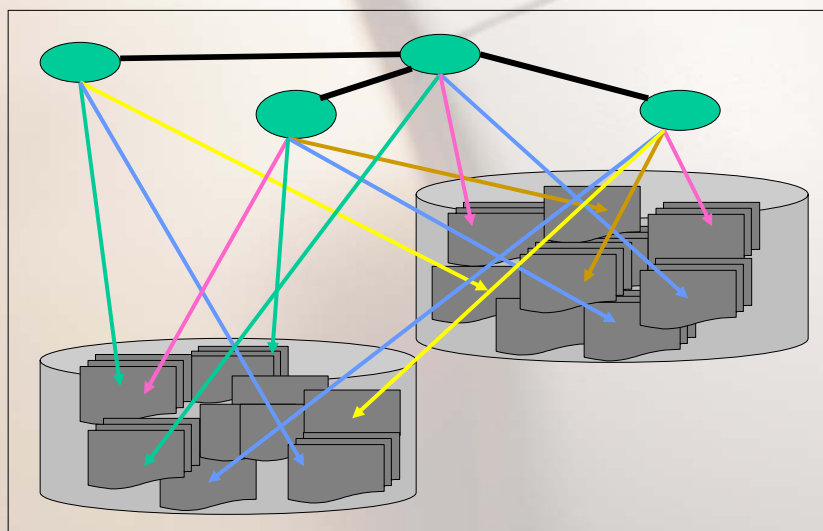
- **Atuais motores de busca = keywords:**
 - Alta retorno de informação, baixa precisão
 - Sensível ao vocabulário
 - Insensível ao conteúdo implícito

- **Exemplo:**
 - Cidades em Portugal onde está a chover neste momento.

A Web nos dias de hoje...



Semantic Web



Ontologia

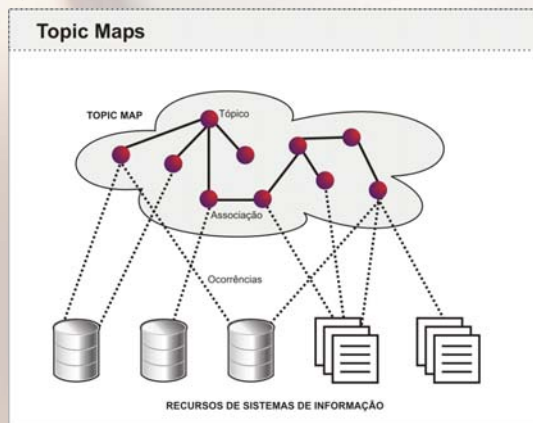
- Segundo o Michaelis: “Ciência do ser em geral.”
- Este termo também é usado em outras áreas, tal como na filosofia, onde tem outro significado.
 - Parte da metafísica que estuda o ser em geral e suas propriedades transcendentais.
- É uma especificação ou formalização de uma contextualização (by Gruber).
- Em TM: É uma precisa descrição dos tipos de elementos que são encontradas no domínio coberto pelo Topic Map.
- Em outras palavras, o conjunto de tópicos que é usado para definir classes de tópicos, associações, papéis de atuação e ocorrências.

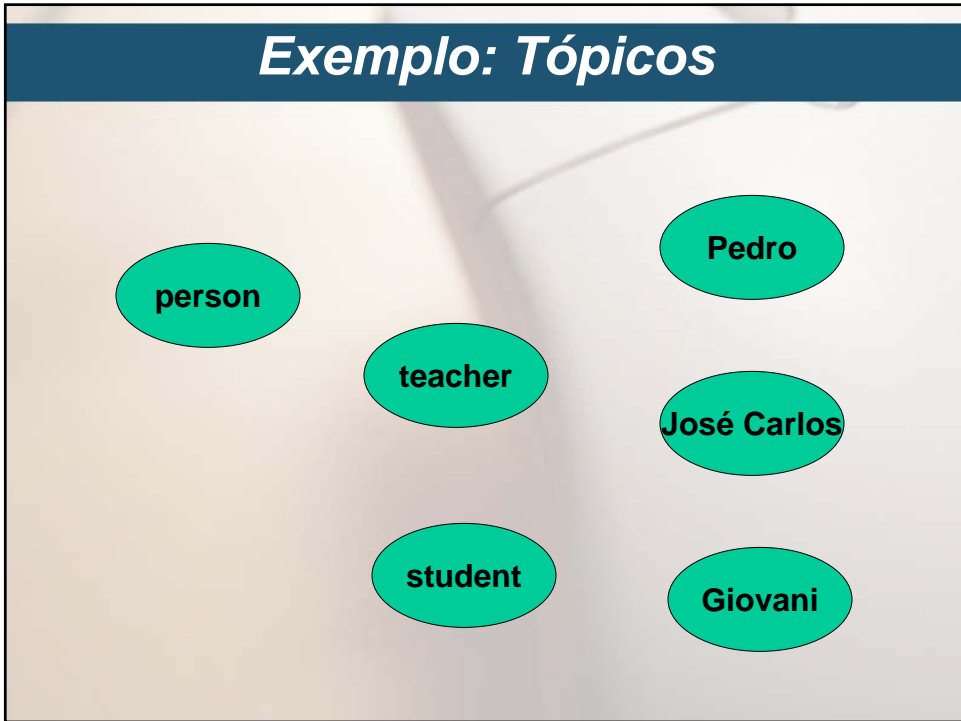
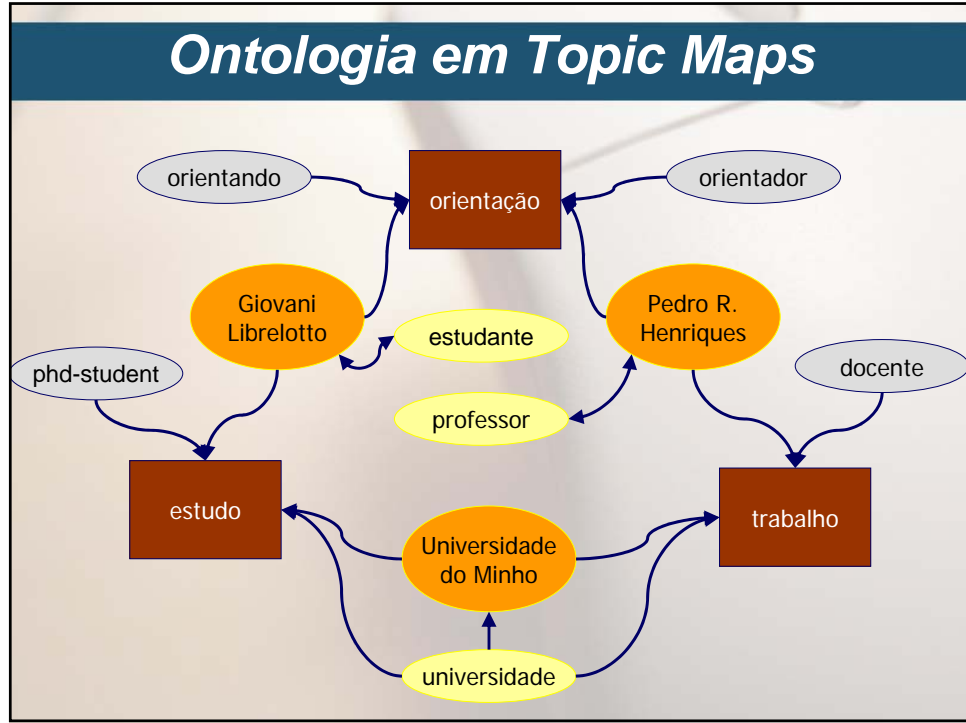
Topic Maps

- **“Topic maps are a new ISO standard for describing knowledge structures and associating them with information resources”**

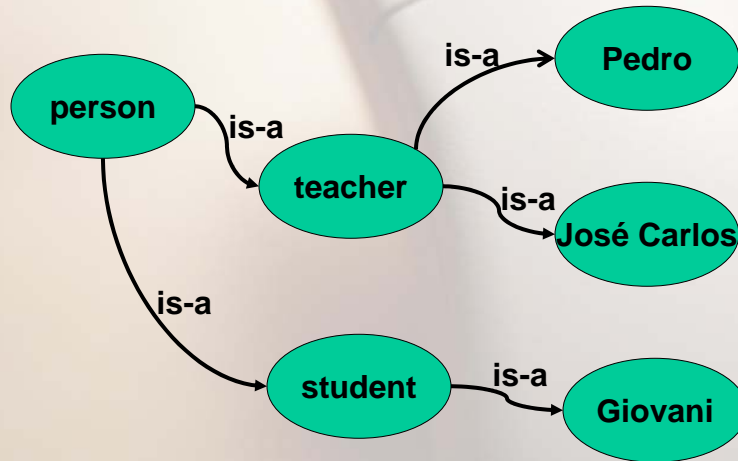
The TAO of Topic Maps,
Steve Pepper, 05-2000

- **Tópicos**
- **Associações**
- **Ocorrências**

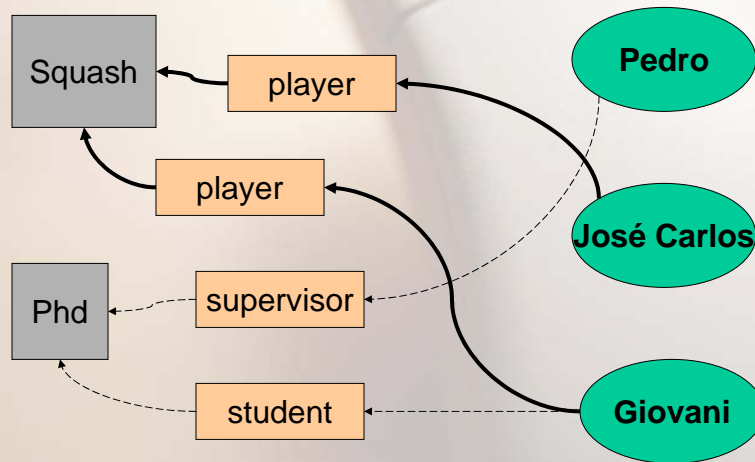




Exemplo: Relações



Exemplo: Associações



Exemplo: Ocorrências

José Carlos

jcr@di.uminho.pt

http://www.di.uminho.pt/~jcr/

+351 253 604479

Índice Back-of-the-book

Tópicos	Múltiplos topic names	Ocorrências
Gorda Sound	see North Sound	
Little Dix Bay		89
North Sound		90
Road Harbour	see also Road Town	73
Road Town		69, 71
Spanish Town		81, 82
<i>Tortola</i>		67
<i>Virgin Gorda</i>		77

Diferentes topic classes

Associações

Diferentes occurrence classes

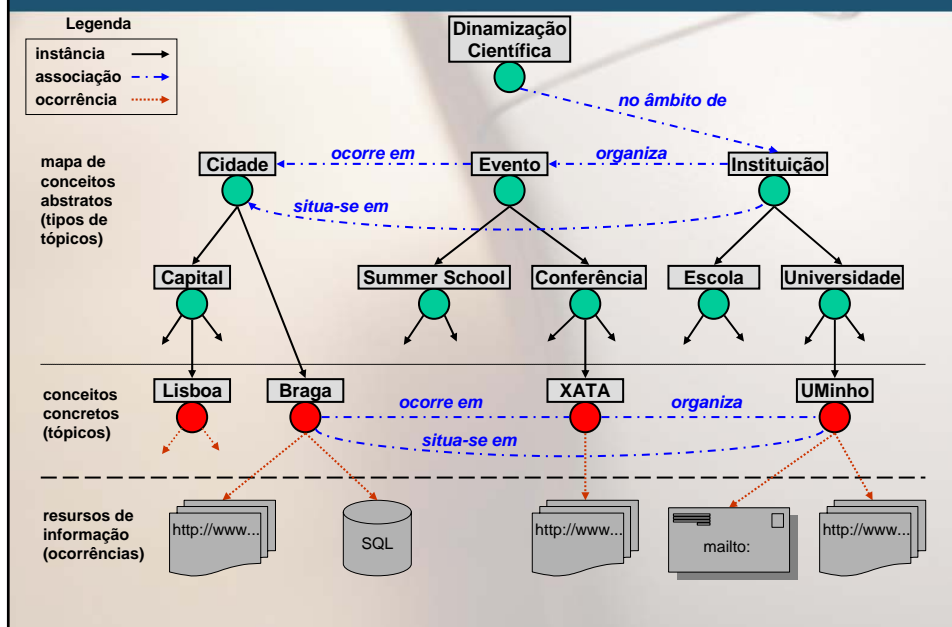
XTM (XML Topic Maps) ...

... é basicamente um documento XML onde diferentes elementos são usados para representar:

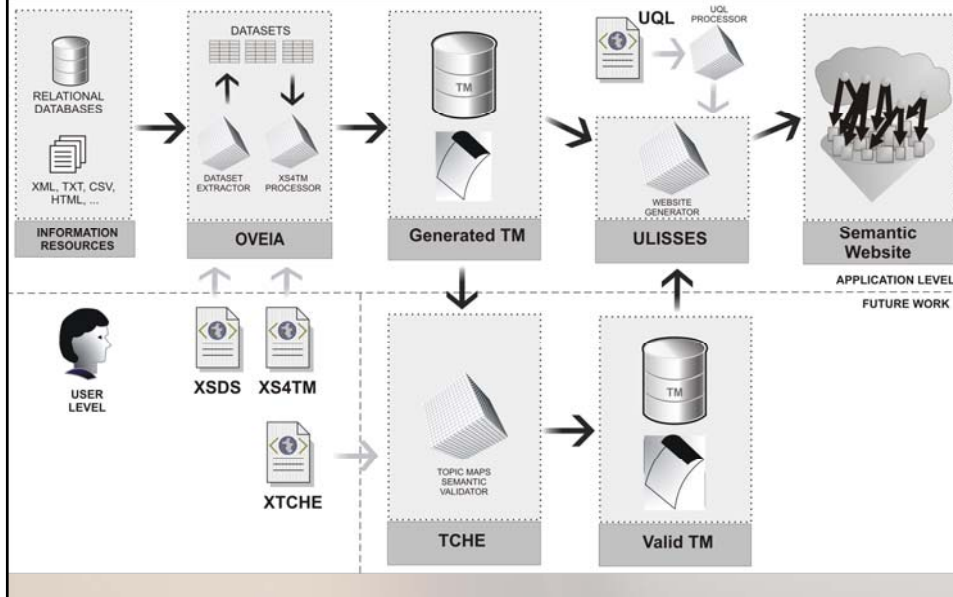
- tópicos,
- ocorrências de tópicos,
- relacionamentos (ou associações) entre tópicos.

- A especificação XTM serve como um formato de intercâmbio XML para Topic Maps.

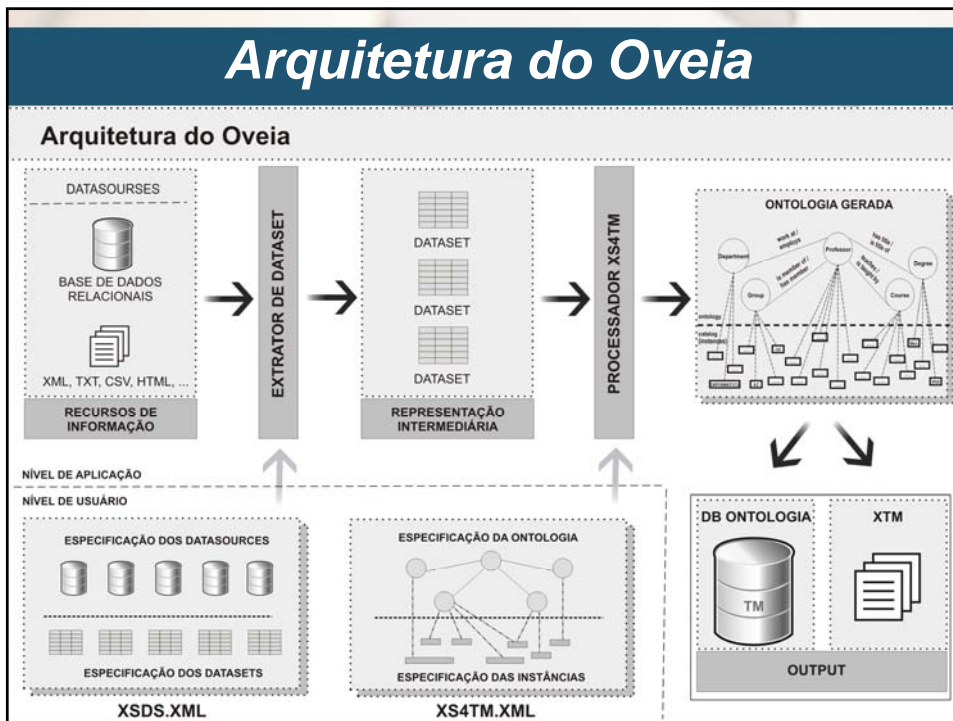
Topic Maps como Mapa Conceptual



Arquitetura do Metamorphosis



Arquitetura do Oveia



XSIDS

- ***XML Specification for Data Sources***
- **Suporta diferentes tipos de fontes (RDBMS, XML, CSV, HTML, ...)**
- **Baseado em drivers:**
 - ex.: Databases e documentos XML**
- **Define a criação de uma representação intermediária (*Dataset*)**

XS4TM

- ***XML Specification for Topic Maps***
- **Especificação da extração de Ontologias**
- **Definido por um especialista do domínio em questão**
- **XTM tornou-se um sub-conjunto de XS4TM**
- **2 partes:**
 - **Ontologias**
 - **Instâncias (catálogo)**

Context Free Grammar de XS4TM

XS4TM	::= Ontologies Instances
Ontologies	::= (Topic Association)*
Instances	::= (Topic Association)*
Topic	::= id (InstanceOf* SubjectIdentity? (BaseName Occurrence)*)
InstanceOf	::= id (TopicRef ResourceRef SubjectIndicatorRef)
SubjectIdentity	::= id (TopicRef ResourceRef SubjectIndicatorRef)*
TopicRef	::= id xlink:type xlink:href
SubjectIndicatorRef	::= id xlink:type xlink:href
BaseName	::= id (InstanceOf? Scope? BaseNameString Variant*)
BaseNameString	::= id baseNameString
Variant	::= id (Parameters VariantName? Variant*)
VariantName	::= id (ResourceRef ResourceData)
Parameters	::= id (TopicRef ResourceRef SubjectIndicatorRef)+
Occurrence	::= id (InstanceOf? Scope? (ResourceRef ResourceData))
ResourceRef	::= id xlink:type xlink:href
ResourceData	::= id resourceData
Association	::= id (InstanceOf? Scope? Member+)
Member	::= id (RoleSpec? (TopicRef ResourceRef SubjectIndicatorRef)*)
RoleSpec	::= id (TopicRef ResourceRef SubjectIndicatorRef)
Scope	::= id (TopicRef ResourceRef SubjectIndicatorRef)+

The screenshot shows two XML documents side-by-side in XMLSpy. The left document, [XSDS.xml], contains a list of datasets. One dataset is highlighted with a red circle:

```
<dataset name="DS_aluno" database="BD_DI_MSSC"
SELECT aluno_id,nome,email,tm, curso_id FROM alur
</dataset>
```

The right document, [XS4TM.xml], contains an ontology structure. A topic is highlighted with a red circle:

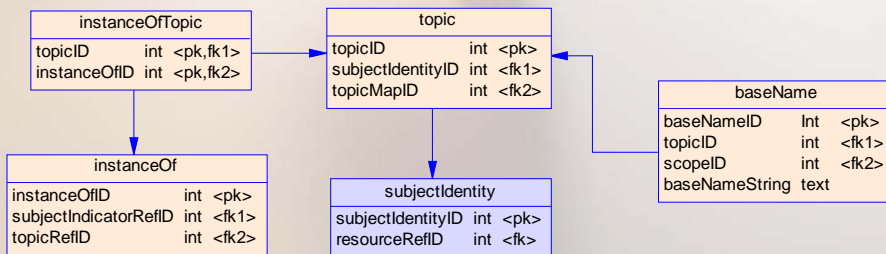
```
<topic dataset="@DS_aluno">
<instanceOf>
<topicRef xlink:href="#aluno"/>
</instanceOf>
<baseName>
<baseNameString>@DS_aluno.nome</baseNameString>
</baseName>
<occurrence>
<instanceOf>
<topicRef xlink:href="#numero"/>
</instanceOf>
<resourceData>@DS_aluno.rm</resourceData>
</occurrence>
<occurrence>
<instanceOf>
<topicRef xlink:href="#email"/>
</instanceOf>
<resourceRef xlink:href="@DS_aluno.email"/>
</occurrence>
</topic>
```

Red arrows indicate the mapping between the two documents. One arrow points from the 'email' resourceRef in the right document to the 'DS_aluno' dataset in the left document. Another arrow points from the 'DS_aluno' dataset in the left document to the 'DS_aluno' topic in the right document.

OntologyDB

- Database modelada de acordo com o paradigma *Topic Maps*

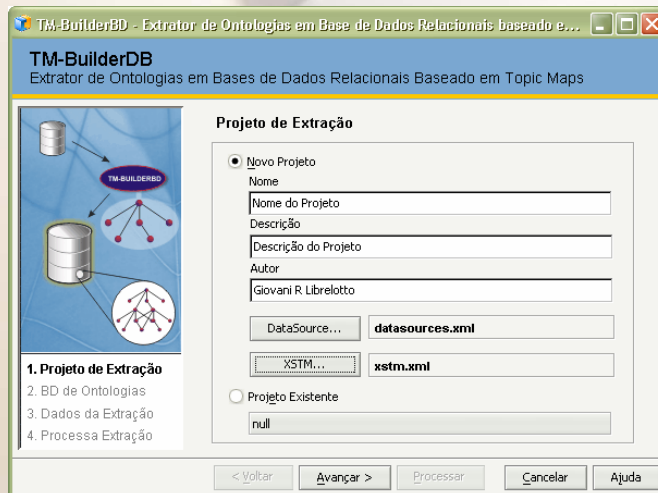
- Suporta todos os elementos de XTM



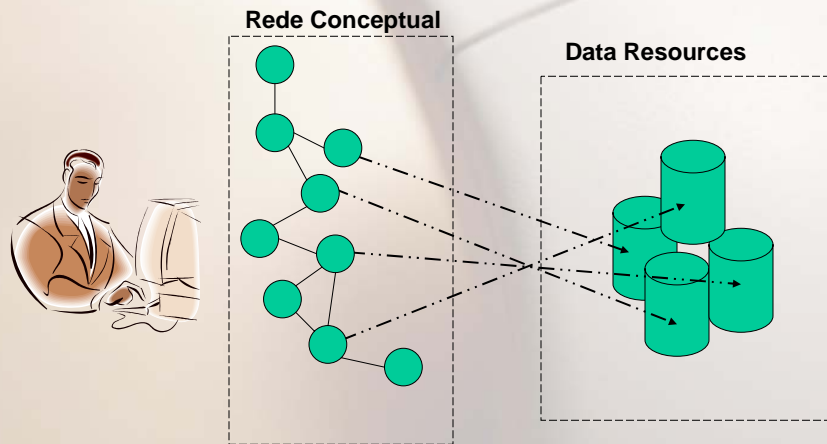
...

Protótipo do Oveia

- Ferramenta de extração visual



Ulisses



Cenários de Utilização

- E-learning
- Proceedings de conferências
- Website departamental
- Biblioteca digital do Arquivo Histórico
- Integração de sistemas de informação
- Outros pequenos casos de estudos

Trabalho Futuro

- Desenvolver uma interface gráfica (SVG)
- Desenvolver uma interface adaptativa/dinâmica
- Front-end para XSDS e XS4TM
- Website do Metamorphosis

Um Extrator de Topic Maps a partir de Recursos Heterogêneos de Informação

Giovani Rubert Librelotto
José Carlos Ramalho
Pedro Rangel Henriques

{grl,jcr,prh}@di.uminho.pt

Universidade do Minho - Braga - Portugal