Topic Maps Extraction in Oveia: Specification and Processing

Extração de Topic Maps no Oveia: Especificação e Processamento

Giovani R. Librelotto
José Carlos Ramalho
Pedro R. Henriques
Department of Informatics
University of Minho
Portugal

Motivation

- Suppose you have an information system with several heterogeneous data resources:
  - Relational databases, XML documents, etc…
- You want to achieve semantic interoperability between those data resources;
- You want to do it fast

Motivation

- The use of ontologies is a good approach to overcome the problem of semantic heterogeneity;
- This supports the usefulness of Topic Maps;
- However tools to build Topic Maps are crucial because the Topic Maps creation is an hard task.

Index

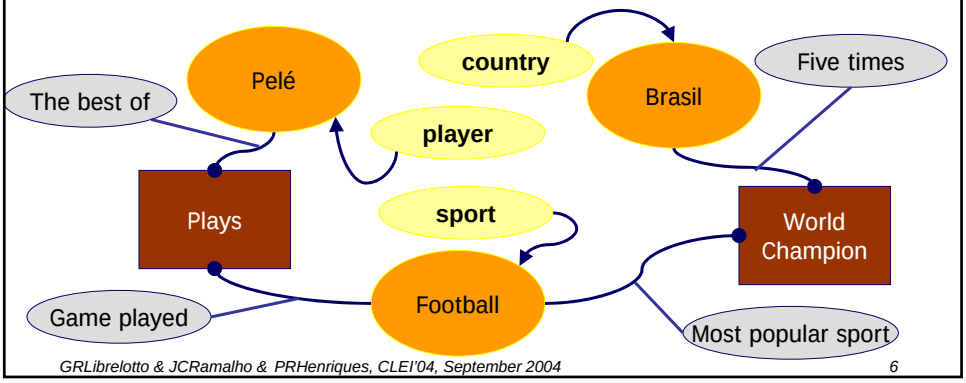
- Basic Concepts
- Our approach
- Inside Oveia
- Case Study
- Conclusion

Ontology

- Metaphysical branch of study which is concerned with existence and the nature of being;

Ontology

- An ontology is just a set of words and relationships that formally describes an universe of discourse or context.

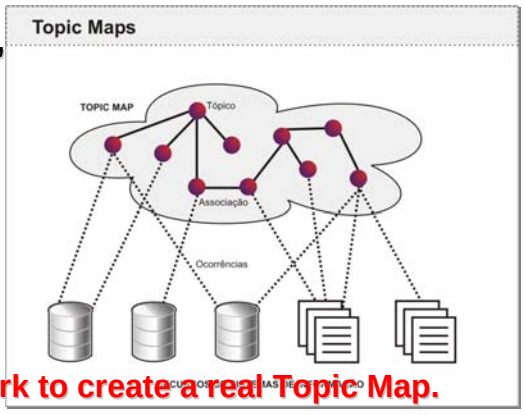


Ontology Specification

- Specifications Standards:
  - RDF(S): Resource Description Format
  - DAML/OIL: Darpa Agent Markup Language
  - OWL: Ontology Web Language
  - XTM: XML Topic Maps (our choice)

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
- ***Topics***
- ***Associations***
- ***Occurrences***
- **However too much work to create a real Topic Map.**



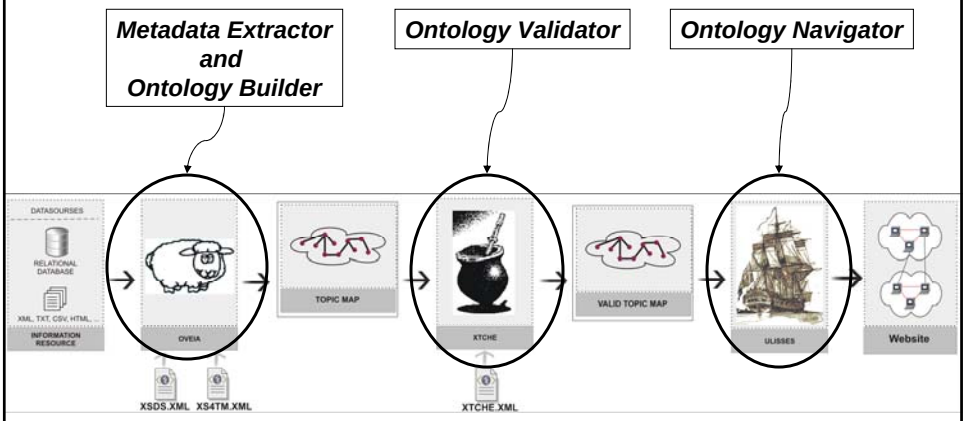
Ontology Support

- 94 tools and similar environments to support creation, use, and maintenance
  - ***Ontology Tools Survey, Revisited***
    by Michael Denny, July 14, 2004, www.xml.com
- However no one for the automatic creation of Topic Maps.

Index

- Basic Concepts
- Our approach
- Inside Oveia
- Case Study
- Conclusion

Metamorphosis



Index

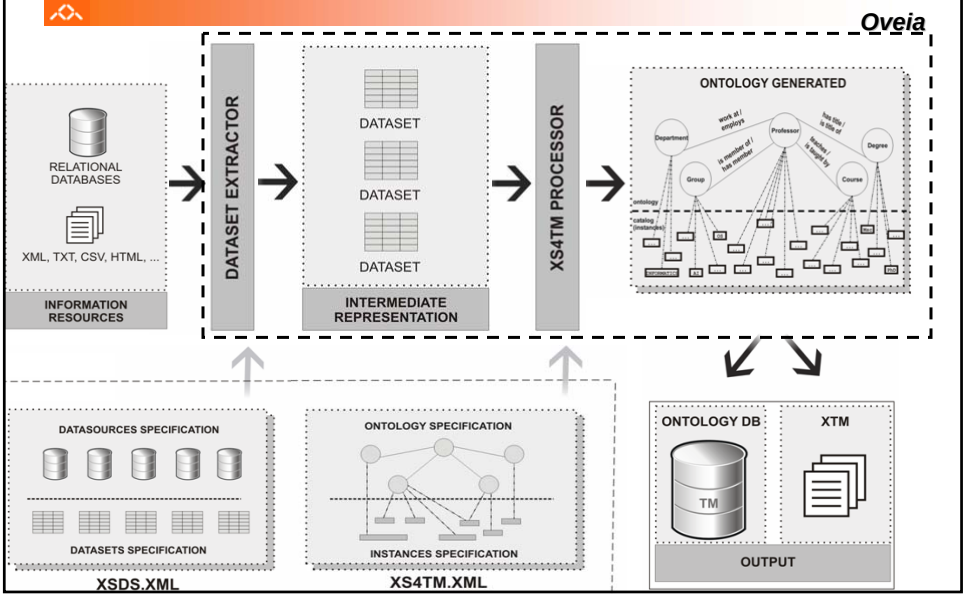
- Basic Concepts
- Our approach
- Inside Oveia
- Case Study
- Conclusion

Oveia

- A Topic Maps extractor from heterogeneous information system composed of two engines:
  - Metadata Extractor: collects pieces of information and stores them in an intermediate representation;
  - Ontology Builder: uses a specification to transform the intermediate representation into an ontology according to Topic Maps standard.

Oveia

Metadata Extractor + Ontology Builder



Metadata Extractor

- XSDS (XML Specification of Data Sources)
- Supports different kinds of sources (relational databases, XML files, …)
- Uses a *driver* for each data source
- Creates an intermediary representation (called *Dataset*)

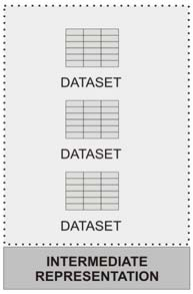
Extractor Specification

```
<resources>
  <datasources>
    <datasource
        extratorDriver="br.uneb.dcet.tmbuilder.drivers.DataBase"
        name="xata2004">
      <parameter  name="connectionURL">
         jdbc:mysql://localhost/XATA2004
      </parameter>
      <parameter name="password"/>
      <parameter name="user">root</parameter>
      <parameter name="jdbcDriver">
         org.gjt.mm.mysql.Driver
      </parame
    </datasou
  </datasourc
  <datasets>
    ...
  </datasets>
</resources>
```

```
<dataset name="Authors" database="xata2004">
      SELECT code, name, url FROM author-table
</dataset>
<dataset name="Papers" database="xata2004">
      SELECT code, title FROM paper-table
</dataset>
```

Datasets

- An intermediary representation;
- Contains all data extracted from information resources;
- Is the input to the *XS4TM processor*;
- Data is stored in table format:
  - Line x collumn



Ontology Builder

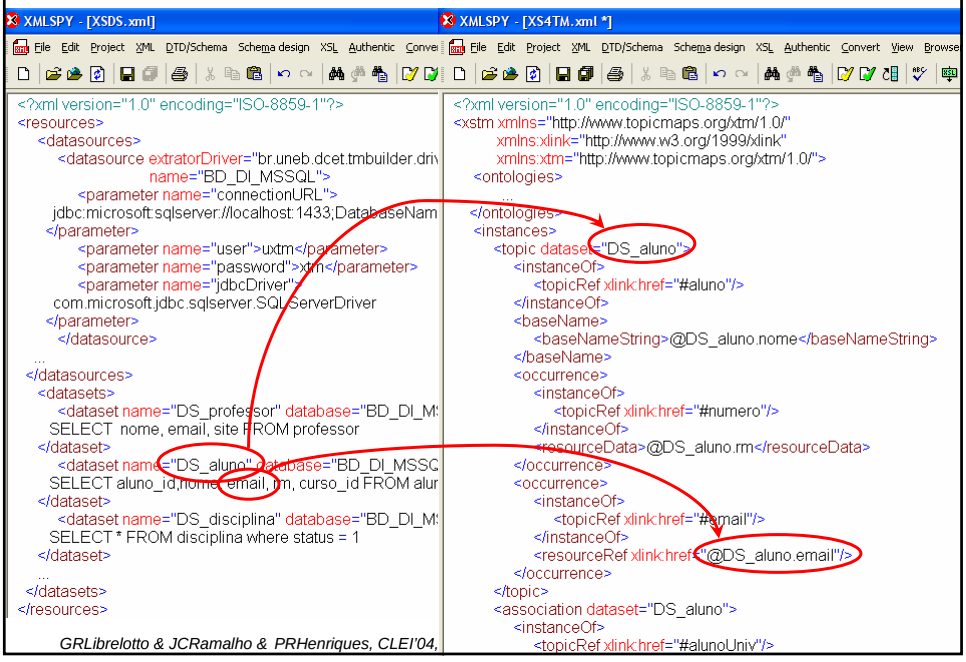
- XS4TM (XML Specification for Topic Maps)
  - Ontology extraction specification
- XTM becomes a sub-set of XS4TM
- XS4TM has 2 parts:
  - Abstract Structure
  - Instances (catalog)

OntoBuilder Specification

```
<instances>
  <topic dataset="Categorias">
    <instanceOf>
      <topicRef xlink:href="#Categorias"/>
    </instanceOf>
    <baseName>
      <baseNameString>
        @Categorias.Descricao
      </baseNameString>
    </baseName>
  </topic>
  ...
</instances>
```

Reference to the extracted dataset

XSDS x XS4TM

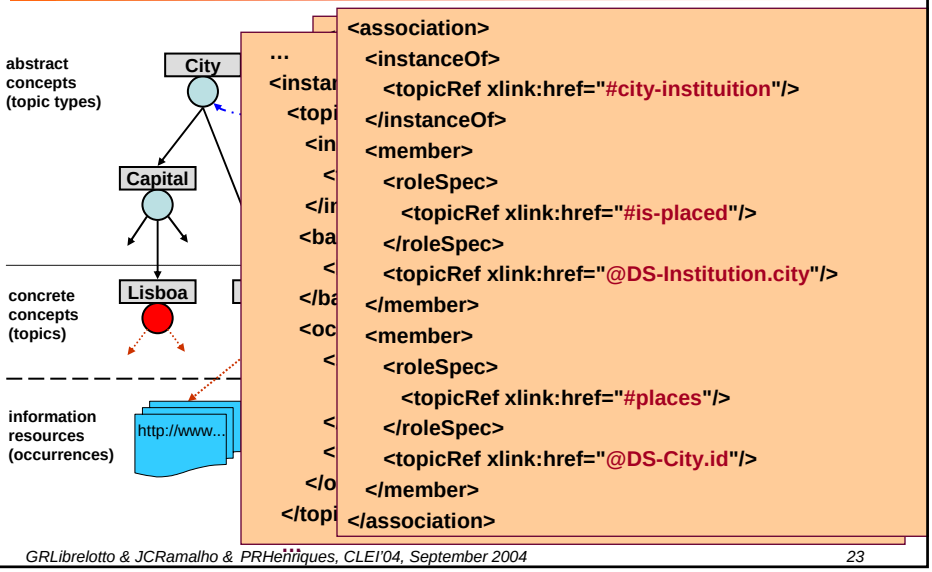


Generated topic map

- After the XS4TM processing, *Oveia* generates a topic map stored in memory;
- *Oveia* has two possible output formats:
  - XTM file: an XML document.
  - OntologyDB: a relational database designed according to *Topic Maps* standard.

Index

Case Study: Conferences



Conclusion

- This presentation appears in the context of the integration of heterogeneous information systems using the ontology paradigm and suggests the use of Topic Maps to describe the ontologies.
- *Oveia* is an architecture for the automatic construction of Topic Maps with data extracted from information systems.

Future Work

- Front-end development:
  - *XSDS*: datasource spec.
  - *XS4TM*: ontology builder spec.

- Part of this work is being integrated in an european **Eureka** project: IKF-P E!2235 "*Information Knowledge Fusion*"