CORE

Metadata, citation and similar papers at core.ac.uk

Provided by Universidade do Minho: RepositoriUM

Extreme Markup Languages 2007®

Montréal, Québec
August 7-10, 2007

Topic Maps applied to PubMed

Giovani Rubert Librelotto
UNIFRA

Mirkos Martins
UNIFRA

José Carlos Ramalho
University of Minho

Henrique Tamiosso Machado
UNIFRA

Pedro Gabriel Dias Ferreira
University of Minho

Pedro Rangel Henriques
University of Minho

Abstract

This paper presents a topic map approach to PubMed in order to create a knowledge representation for this information system. PubMed is a free search engine that gives very full coverage of the related biomedical sciences. With more than 17 millions of citations since 1865, PubMed users have several problems to find the papers desired. So, it is necessary to organize these concepts in a semantic network. To achieve this objective, we use the Metamorphosis system, choosing the keywords from MeSH ontology. This way, we obtain an ontological index for PubMed, making easier to find specific papers.

Extreme Markup Languages®

Topic Maps applied to PubMed

Table of Contents

Introduction....1

PubMed....1

Metamorphosis....2

- Metamorphosis Repository....4
- Topic Map Discovery....5
- Topic Map Extractor....6
- Topic Map Validator....7

Topic Maps applied to PubMed....7

- Defining the Topic Maps concepts to PubMed citations....9
- Browsing the topic map....9

Conclusion....10

Bibliography....12

The Authors....13

Topic Maps applied to PubMed

Giovani Rubert Librelotto, Henrique Tamiosso Machado, Mirkos Martins, Pedro Gabriel Dias Ferreira, José Carlos Ramalho, and Pedro Rangel Henriques

§ Introduction

Daily, a lot of data is stored into PubMed system. There is a problem that organization requires an integrated view of their heterogeneous information systems. In this situation, there is a need for an approach that extracts the information from their data sources and fuses it in a semantically network. Usually this is achieved either by extracting data and loading it into a central repository that does the integration before analysis, or by merging the information extracted separately from each resource into a central knowledge base.

Topic maps are an ISO standard for the representation and interchange of knowledge, with an emphasis on the findability of information. A topic map can represent information using topics (representing any concept), associations (which represent the relationships between them), and occurrences (which represent relationships between topics and information resources relevant to them). They are thus similar to semantic networks and both concept and mind maps in many respects. According to Topic Map Data Model (TMDM) [GM05], Topic Maps are abstract structures that can encode knowledge and connect this encoded knowledge to relevant information resources. In order to cope with a broad range of scenarios, a topic is a very wide concept. This makes Topic Maps a convenient model for knowledge representation.

This paper described the integration of data from PubMed information system using the ontology paradigm, in order to generate an homogeneous view of this resources. PubMed is introduced in section 2. The proposal uses an environment, called Metamorphosis (section 3), for the automatic construction of Topic Maps with data extracted from the various data sources, and a semantic browser to navigate among the information resources. It is described in section 4. The section 5) presents the concluding remarks.

§ PubMed

PubMed [NLMb] is a free search engine that provides very full coverage of the related biomedical sciences, such as biochemistry and cell biology. It also offers access to the MEDLINE database [NLMa] with citations and abstracts of biomedical research articles.

The PubMed core subject is medicine and its related fields. It is offered by the United States National Library of Medicine as part of the Entrez information retrieval system. The inclusion of an article in PubMed does not endorse the article's contents, as other indexes. Nevertheless, many PubMed citations contain links to full text articles which are freely available, often in the PubMed Central digital library.

MEDLINE database covers over 4.900 journals published around the world primarily from 1966 to the present and is composed of more than 17 millions of citations. Information about the journals indexed in PubMed is found in its Journals Database, searchable by subject or journal title, Title Abbreviation, the NLM ID (NLM's unique journal identifier), the ISO abbreviation, and both the print and electronic International Standard Serial Numbers (pISSN and eISSN). The database includes all journals in all Entrez databases. A PubMed entry includes among other information the following details: PubMed identifier, Authors' name, Title, Journal, Publication date, Language, and Mesh terms.

The PubMed database consists of three tiers of software as shown in Figure 1. At the bottom is a database management system (DBMS) that manages a collection of facts. At the top is the web browser that transmits requests for data to the database and renders the responses as web pages. In the middle is a software layer that mediates between the DBMS and the web browser to turn data requests into database queries, and to transform the query responses into hypertext mark-up language (HTML).

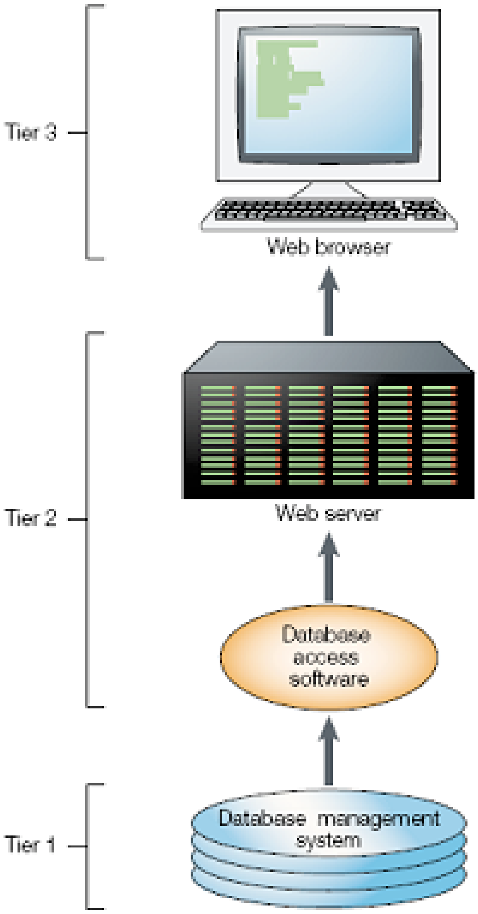
The PubMed data structure is composed of citations metadata. Each citation has the same structure. The main part of its schema can be formalized by the following context free grammar:

```
MedlineCitation ==> PMID, DateCreated, DateCompleted, Article,
MedlineJournalInfo, ChemicalList,
CitationSubset, MeshHeadingList
Article ==> Journal, ArticleTitle, Pagination,
```

Copyright © 2007 by the author. Reproduced with permission of the author.

```
Abstract, Affiliation, AuthorList,
Language, PublicationTypeList
Journal ==> ISSN, JournalIssue, Title
JournalIssue ==> Volume, Issue, PubDate
PubDate ==> Year, Month, Day, Hour?, Minute?, Second?
MedlineJournalInfo ==> Country, MedlineTA, NlmUniqueID
ChemicalList ==> Chemical+
```

Figure 1: PubMed database architecture



PubMed files are intended for automatic processing and therefore available in XML format. Each set of 30.000 PubMed citations is stored as an XML instance defined by a DTD. Notice that the context free grammar above was obtained direct and systematically from the PubMed DTD. For these reasons, it was defined an XML Schema to PubMed files. The view of this structure is shown in figure 2.

§ Metamorphosis

The main idea behind Metamorphosis is close the gap between Topic Map technology and its users. Metamorphosis is being developed to become a Topic Map workbench easy to use and accessible to a common user. Figure 3 shows the usage scenario proposed in this paper. It illustrates some of the interaction between the system components, information resources and users.

Metamorphosis can be used to prototype web interfaces or to expose information systems on the web. To do this the user only needs to specify a topic map for each view he wants. Information integration is accomplished by concept integration in the topic map: to integrate two information systems we need to specify the two sets of concepts in the same topic map and specify the associations that will materialize that integration.

Figure 2: PubMed's XML Schema

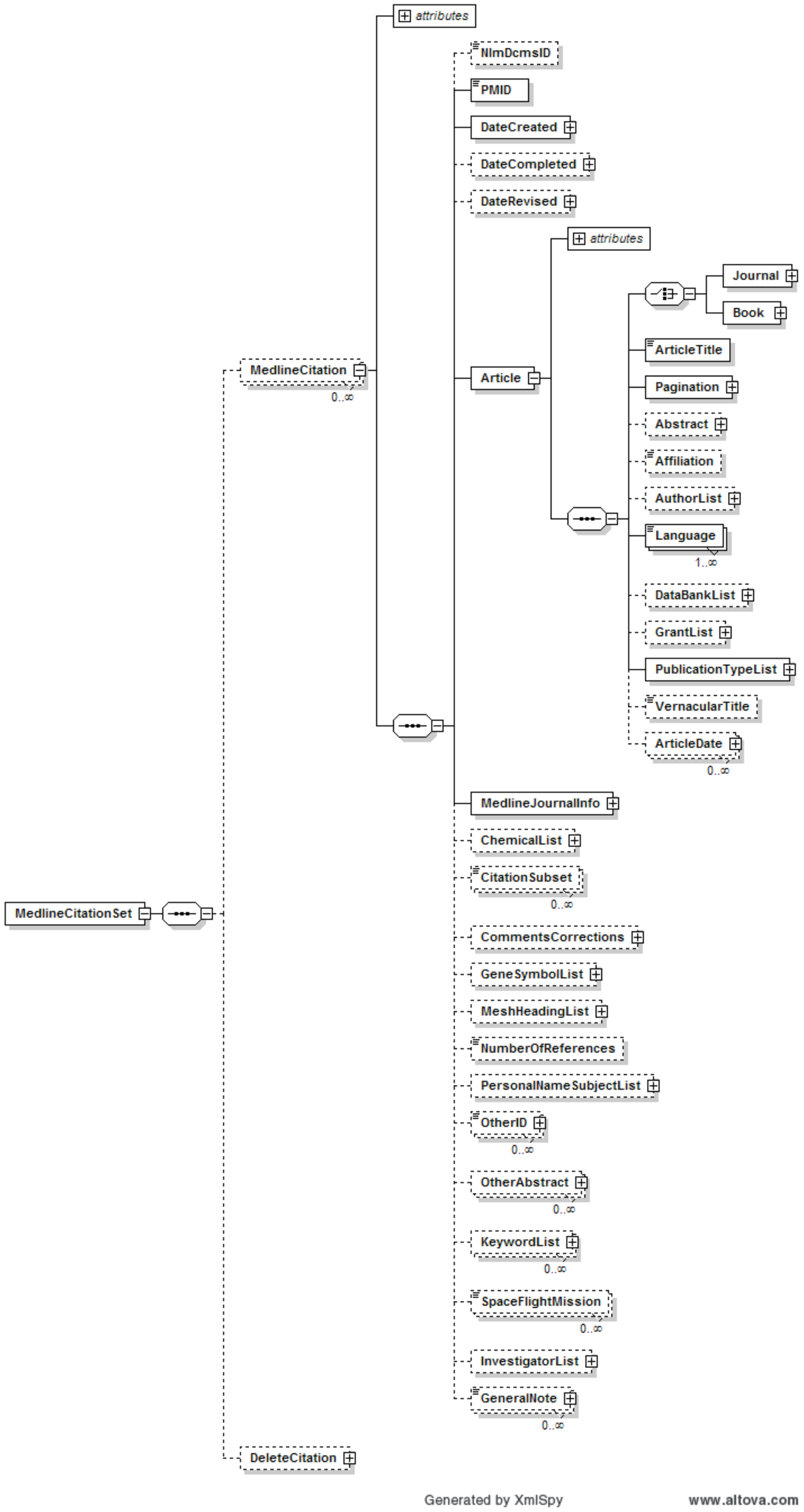
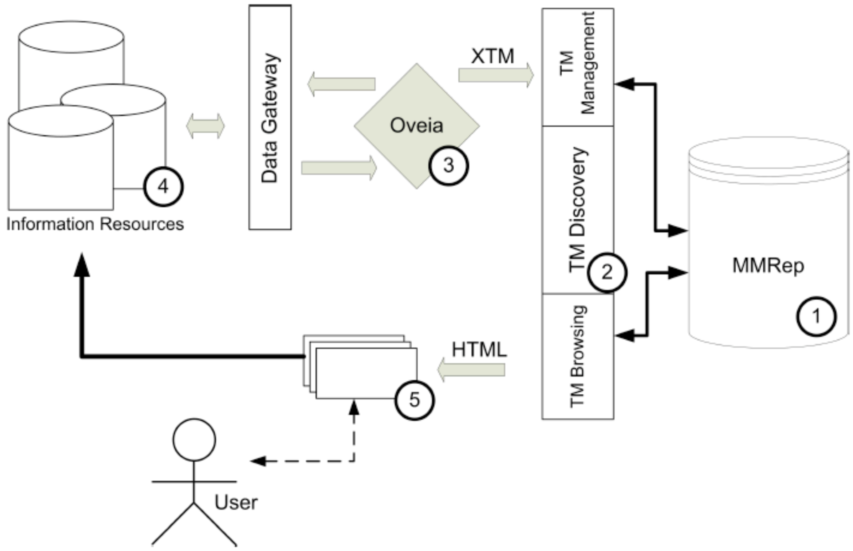


Figure 3: Metamorphosis Functional Diagram



1. Metamorphosis Repository (MMRep) is the central component that takes care of Topic Map storage and management. All the other components interact with MMRep.
2. Topic Map Discovery (TMDiscovery) is a Topic Map driven browser that allows users to navigate inside the Topic Maps stored in MMRep.
3. Topic Map Extractor (Oveia) automates the task of Topic Map harvesting; it enables the user to specify the extraction task and generates a Topic Map in XTM syntax that can be uploaded into MMRep. Oveia implements some extraction mechanisms with which is possible to populate an ontology.
4. Information resources that we want to access.
5. Web interface driven by a topic map stored in MMRep that provides access to information resources.

In the next sections we are going to discuss the main components of this workbench prototype: Metamorphosis Repository, Topic Map Discovery, Oveia and XTche.

This way, Metamorphosis let us achieve the semantic interoperability among heterogeneous information systems because the relevant data, according to the desired information specified through an ontology, is extracted and stored in a topic map. The environment validates this generated topic map against a set of rules defined in a constraint language. That topic map provides information fragments (the data itself) linked by specific relations to concepts at different levels of abstraction. Note that not all data items need to be extracted from the sources to the Topic Map. We only extract the necessary metadata to build the intended ontology. This ontology will have links to enable a browser to access all data items.

Thus the navigation over the topic map is led by a semantic network and provides an homogeneous view over the resources - this justifies our decision of call it semantic interoperability.

Metamorphosis Repository

Although XTM is a good format for interchange it is not so good for storage. When we refer to storage we are meaning the capability of storing a Topic Map and efficiently being able to query it. XTM is easy to process and for instance to translate it into another format. But querying XTM is complex.

The Topic Map model is not hierarchical, every relation is materialized as a reference. Gathering all the information about a topic is very complex. The obvious choice for storage is a database. For this case we had three options: an XML database [Bou05], an Object Oriented Database [Lea00] or a Relational Database. Since the Topic Map model does not match the XML model XML databases were discarded.

Almost for the same reasons OO databases were also discarded. That left us with the relational model as the target for our storage solution.

The next step would be the specification of a Topic Map Relational Model. We have considered two approaches: look at the Topic Map ReferenceModel [GM05] [Kip03] and derive the relational model from it or look at the XTM model and work from there.We decided to work over the XTM model and see if we could reach a model similar to the Topic Map Reference Model.

Data Model: First, we looked at the XTM model and raised the following subject list (and correspondent content model):

```
topicMap = (topic|association|mergeMap)*
topic = (instanceOf|subjectIdentity|baseName|occurrence)*
instanceOf = (topicRef|subjectIndicatorRef)
subjectIdentity = resourceRef|(topicRef|subjectIndicatorRef)*
baseName = (scope?|(topicRef|subjectIndicatorRef|resourceRef)+|baseNameString|
variant*)
scope = (topicRef|subjectIndicatorRef|resourceRef)+
variant = (parameters, variantName?, variant*)
parameters = (topicRef|subjectIndicatorRef)+
variantName = (resourceRef|resourceData)
occurrence = (instanceOf?, scope?, (resourceRef|resourceData))
scope = (topicRef|subjectIndicatorRef|resourceRef)+
association = (instanceOf?, scope?, member+)
member = (roleSpec?, (topicRef|subjectIndicatorRef|resourceRef)*)
mergeMap = (topicRef|subjectIndicatorRef|resourceRef)*
```

After some exercise with the leaf nodes of this list we end with the following types that cover any element in a topic map:

Table 1

(topicRef_subjectIndicatorRef_resourceRef)
(topicRef_subjectIndicatorRef)
(resourceRef_resourceData)
resourceRef
baseNameString

This result means that any Topic Map node can be represented with one of this five types. To store any of this five types we only need a triple: identifier, value and type. Consider the following example in 2.

Table 2: Stored Values

Id	Type	Value
TR982	topicRef	#University
SIR500	subjectIndicatorRef	http://www.uminho.pt
BNS32	baseNameString	U. Minho
RD444	resourceData	UM is ...
RD446	resourceRef	http://www.uminho.pt/students

This exercise enabled us to simplify the model and to reach the relational model showing in Figure 4.

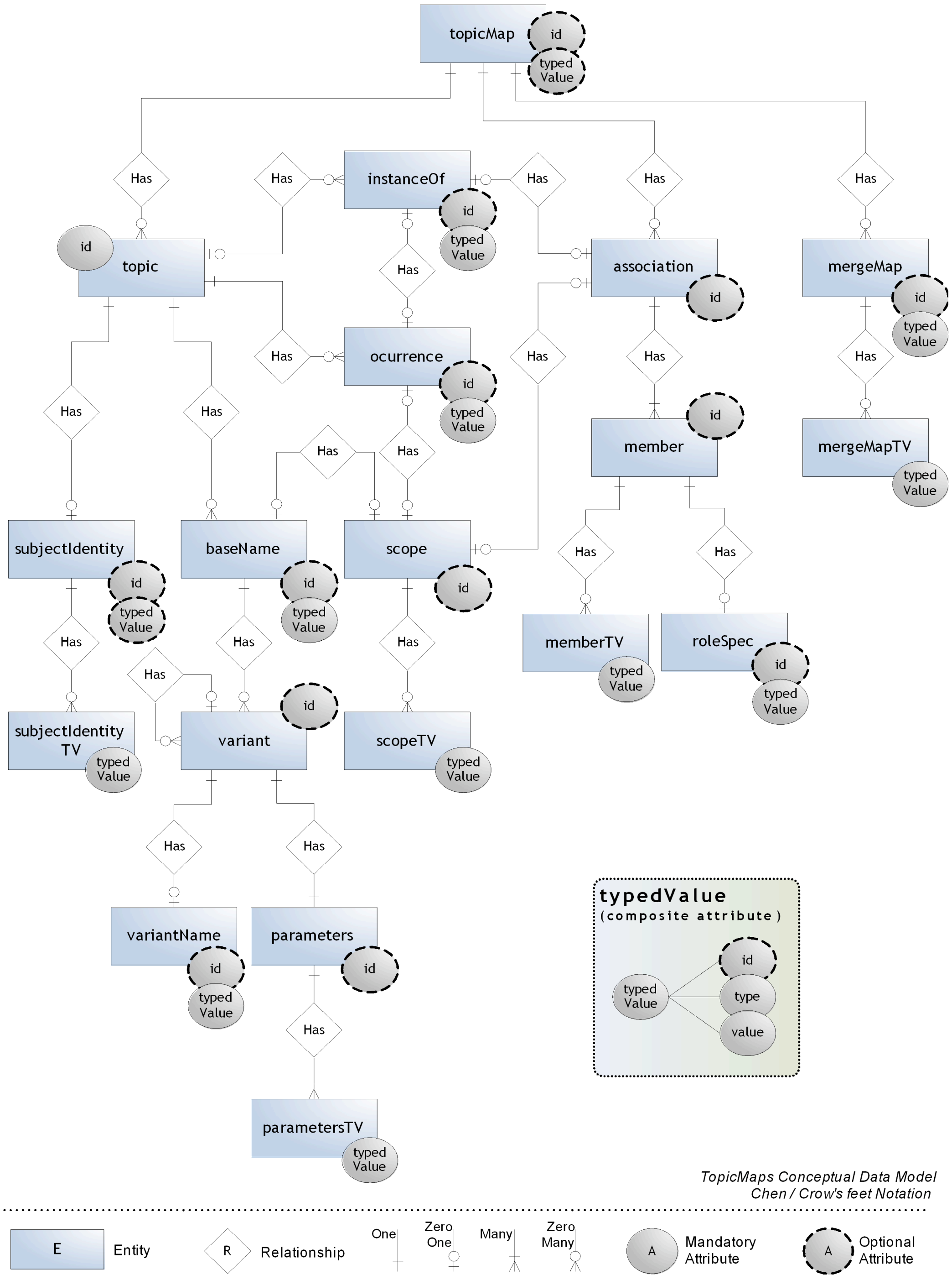
With this specification we have implemented a Topic Map Repository that is the core component of Metamorphosis. In the following sections we will give some details about the integration of the other components with the repository.

Topic Map Discovery

Topic Map Discovery is an API that is being developed in order to work with the repository. For the moment it is composed of two parts: a topic map manager and a browser.

The topic map manager lets you upload and download topic maps in XTM syntax and delete a topic map from the repository (soon it will enable the user to edit stored topic maps).

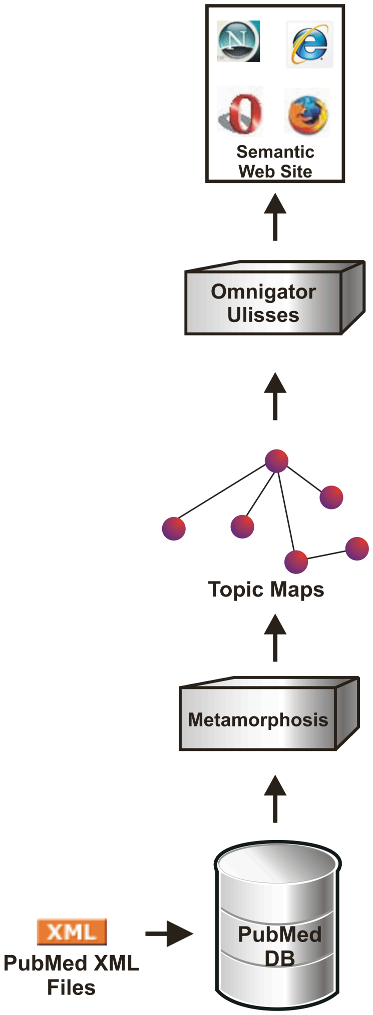
Figure 4: Relational model schema



Topic Map Extractor

The ontology extractor Oveia is based on ISO/IEC 13250 Topic Maps [BBN99].

Figure 5: The system's architecture



Oveia extracts information fragments from heterogeneous information systems according to an XSDS specification and builds the topic map according to an ontology specified in XS4TM language. More details in [LRH04a] [LSRH04].

Topic Map Validator

XTche is a specification language that allow us to define the schema of a Topic Maps family. This language meets all the requirements established by ISO Working Group for a Topic Map Constraint Language (TMCL) [NMB04]. XTche is designed to allow users to constrain any aspect of a topic map. More details in [RLH05] [LRH04b]

§ Topic Maps applied to PubMed

In order to obtain a semantic network from PubMed data, we divided this task in a few parts, as shown in Figure 5.

In the first one, we created a relational database to store all contents of XML data obtained from PubMed data source.This database is generated according to the PubMed DTD using the Exult tool. An SQL script processes the result database to remove the redundant data and to erase several tables unnecessary. The final PubMed local database has 57 tables.

To extract data from this database we use Metamorphosis [LRH06]. Metamorphosis has mechanisms to query the PubMed local database (Oveia) according to an ontology specification (XS4TM). Besides, there is a Web interface to make a query over the database. This interface has a text field to the user puts his query. After the query submission, Metamorphosis processes this string finding MeSH terms that describes the desired publications. These terms are structured in a RDF file [AMMS07]. Using these MeSH terms as keywords, Metamorphosis searches articles that match with the user's query.

This search processes includes several fields, like article's title, abstract, keywords, chemical substances, and MeSH terms. When an article satisfies the query, it will be mapped to a topic, as well its main fields, creating associations between them.When the system receive a request, the required data will be collected from the selected databases at runtime. Then it will be further processed and converted into semantically relevant data by Metamorphosis. This resulting data has the standard XTM format. So, one of the advantages of this approach is that no new database is created and no redundant data is produced.

After end of the process, Metamorphosis has all topics and associations stored in its repository. The generated XTM documents can be then processed and displayed to the user by the presentation tier. This way, any topic maps navigator tool is able to browse the semantic network composed by these concepts. For instance, Ulisses [LRH06] allows the topic maps navigation over Metamorphosis' repository and XTM files (in last case, it is also possible to use Ontopia Omnigator [Omni02]). Information is interconnected within a huge knowledge network navigable in any direction.

Defining the Topic Maps concepts to PubMed citations

In order to define the topic map extraction from PubMed instances, the first task is to specify the main concepts (topic types). This way, the topic types in this domain are:

1. **Article:** each article is stored in a tag called <MedlineCitation>;
2. **Author:** the article authors are declared in <Author>;
3. **Keyword:** the keywords are MeSH terms. They are defined in <MeshHeading>;
4. **Publication year:** this metadata is in //PubDate/Year path;
5. **Journal:** all journals are found in <Journal> tag;
6. **Language:** the paper's language are define in <Language>;
7. **Chemical substances:** all chemical items cited in each paper is referenced in <Chemical>;

After the topics choice, the next step is the topic characteristics definition. Below we have the main ones:

a. **Article:** PMID (PubMed identifier), title, pagination, abstract, DOI, ...
b. **Author:** initials, last name, middle name, and first name;
c. **Keyword:** descriptor and qualifier terms;
d. **Journal:** ISSN, title, abbreviation, volume, issue, and publication date;
e. **Chemical substances:** register number and substance name;

At this moment, all topics and its characteristics are defined. The final topic map definition step is the association type specification. The main association types and some roles are described below:

i. Author writes article;
ii. Keyword describes article;
iii. Article is published in a journal;
iv. Article was published in an year;
v. Article is written in a language;
vi. Article refers to chemical substances;
vii. Author publishes in an year;
viii. Author writes paper in a language;
ix. Journal refers to the keywords;

Looking at a TM we can think of it as having two distinct parts: an ontology and an object catalog. The ontology is defined by what we have been designating as topic type, association type, and association role. The catalog is composed by a set of information objects that are present in information resources (one object can have multiples occurrences in the information resource) and that are linked to the ontology.

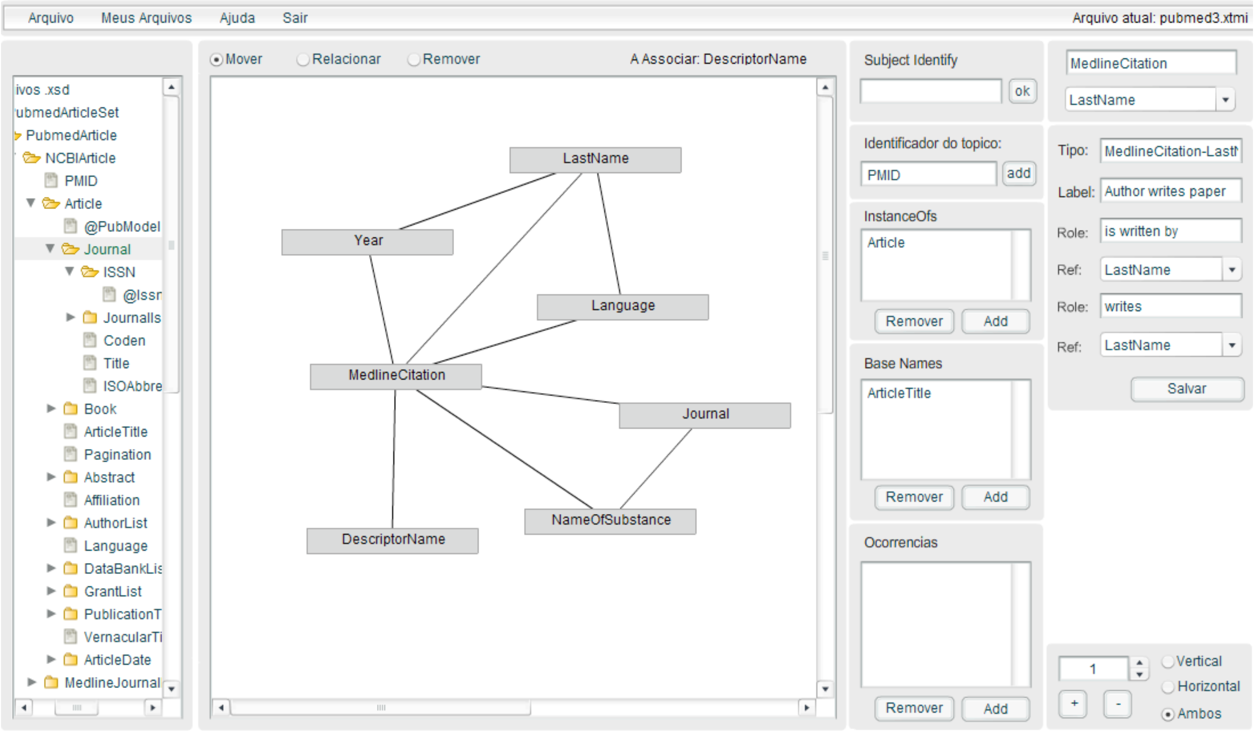
PubMed's topic map ontology defined above (topic types, roles, and association types) and the topic characteristics are mapped to an XS4TM specification as can be seen in next subsection. The XS4TM specification describing the PubMed scenario was defined in a XS4TM Web editor. Figure 6 shows a view of this specification, which defines seven topic types, nine association types, and eighteen role types.

On the left side, XS4TM presents the XML tree extracted from PubMed's XML Schema. The topic types from this case study are shown in the center window. To create a new topic type, the user just needs to make a simple drag and drop from XML tree. The topic characteristics are defined in the first column and the association characteristics are defined in the last column.

With the complete XS4TM specification, Oveia\footnote{Oveia is a Metamorphosis' module} can processes it. Its behavior can be described in four steps: (1) reads the XS4TM specification, (2) extracts the topics and associations from the query result set, (3) creates the topic map, and (4) stores it in the repository.

Browsing the topic map

Figure 6: PubMed's XS4TM Specification



Arquivo atual: pubmed3.xtmi

● Mover ○ Relacionar ○ Remover

A Associar: DescriptorName

Subject Identify:

Identificador do topico: PMID

InstanceOfs: Article

Base Names: ArticleTitle

Ocorrencias:

MedlineCitation

LastName

Tipo: MedlineCitation-Lastn

Label: Author writes paper

Role: is written by

Ref: LastName

Role: writes

Ref: LastName

1

○ Vertical
○ Horizontal
● Ambos

When it will be browsing the semantic network obtained from PubMed local database, Ulisses gives the user an interfaceto navigate inside any of the stored topic maps. It allows the following interfaces:

1. **Topic Maps :**Topic Maps : is the browser entry point and shows a list of all stored topic maps.
2. **Ontology Index :** gives you a structured view of a topic map showing the abstract concepts: topic types, association types, occurrence types, and association role types.
3. **Individuals Index :** lists all non-type topics in alphabetical order.
4. **Full Index :** lists all named topics.
5. **Topic View :** lists a subset of the available information about a topic; for the moment: the basenames, its type, all the associations it participates in together with the other members and their roles, internal occurrences and external occurrences.
6. **Association View :** lists the names associated with the association and all its descendants.

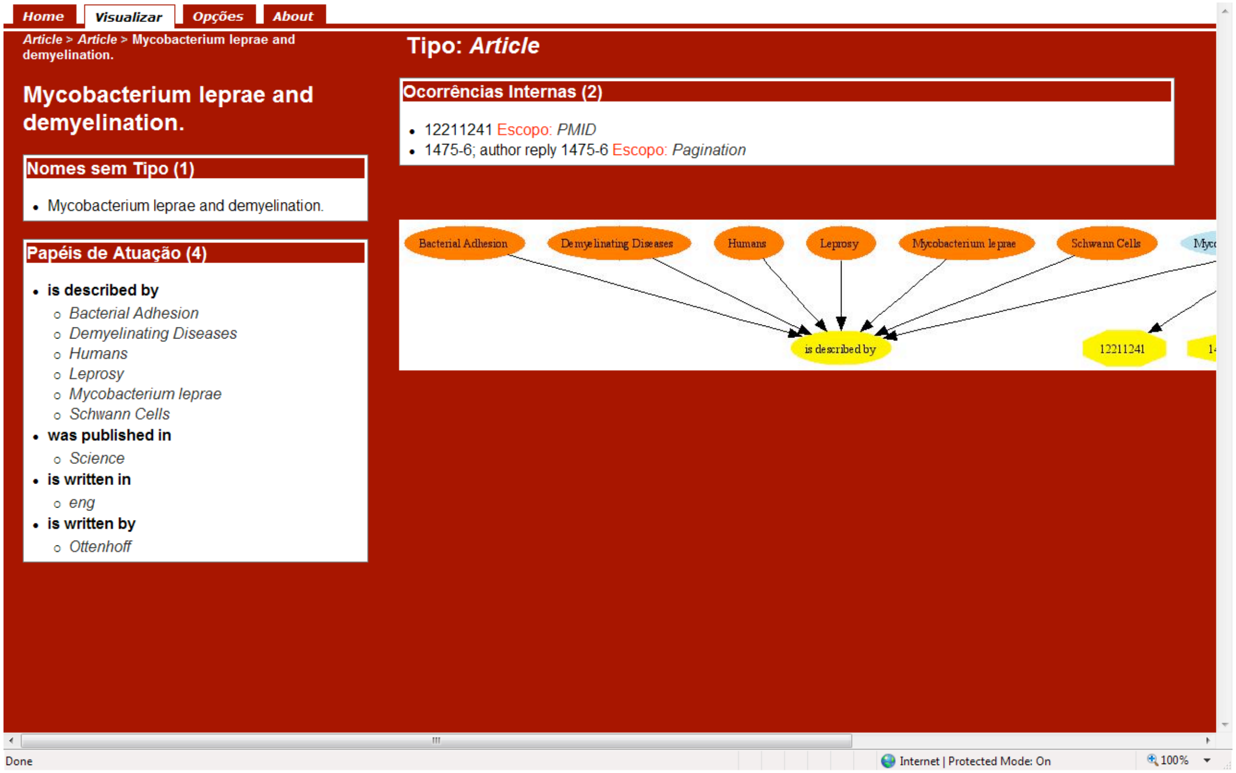
Figure 7 shows a view to the topic of type *article* called *Mycobacterium leprae anddemyelination*. This page display every topic characteristics and its associations in a Web way, as well in a graph view.

Creating a virtual map of the information, Ulisses enables to keep the information systems in their original form, without changes. It is also possible to create as many virtual maps as the user wants generating multiple semantic views for the same sources.

§ Conclusion

This paper described the integration of data from PubMed information system using the ontology paradigm, in order to generate an homogeneous view of this resources. PubMed is a searchable compendium of biological literature that is maintained by the National Center for Biotechnology Information (NCBI).

Figure 7: Ulisses topic view



Article > *Article* > Mycobacterium leprae and demyelination.

Tipo: *Article*

Mycobacterium leprae and demyelination.

Nomes sem Tipo (1)

- Mycobacterium leprae and demyelination.

Papéis de Atuação (4)

- is described by
  - *Bacterial Adhesion*
  - *Demyelinating Diseases*
  - *Humans*
  - *Leprosy*
  - *Mycobacterium leprae*
  - *Schwann Cells*
- was published in
  - *Science*
- is written in
  - *eng*
- is written by
  - *Ottenhoff*

Ocorrências Internas (2)

- 12211241 Escopo: *PMID*
- 1475-6; author reply 1475-6 Escopo: *Pagination*

The proposal uses Metamorphosis for the automatic construction of Topic Maps with data extracted from the various data sources, and a semantic browser to navigate among the information resources.

Topic Maps are a good solution to organize concepts, and the relationships between those concepts, because they follow a standard notation - ISO/IEC 13250 - for interchangeable knowledge representation. In this paper we claimed that the semantic integration of PubMed documents is possible to achieve with Metamorphosis.

In order to achieve this we proposed the following methodology:

1. Look at the information resources and decide how your conceptual view should look like;
2. Choose what information bits must be extracted in order to produce that conceptual view;
3. Specify the extraction task using Oveia;
4. Upload the generated Topic Map into MMRep;
5. Browse it with TMDiscovery and use this interface to access the information resources.

With this methodology the original information resources are kept unchanged and we can have as many different interfaces to access it as we want. We just have to create/generate/specify a Topic Map for each one.

As a future work we aim the integration of Topic Maps and MeSH headings minimizing *false hits* and saving time in the searches. Another project is to identify other useful -- but frequently overlooked -- features of the PubMed database.

Bibliography

[AMMS07] Mark van Assem, Véronique Malaisé, Alistair Miles, and Guus Schreiber. *A Method to Convert Thesauri to SKOS* http://thesauri.cs.vu.nl/eswc06/, 2007

[BBN99] Michel Biezunsky, Martin Bryan, and Steve Newcomb. *ISO/IEC 13250 - Topic Maps. ISO/IEC JTC 1/SC34,*, December 1999. **http://www.y12.doe.gov/sgml/sc34/document/0129.pdf**.

[Bou05] Ronald Bourret. *XML and Databases*, 2005 **http://www.rpbourret.com/xml/XMLAndDatabases.htm**.

[GM05] Lars Marius Garshol and Graham Moore. *Topic Maps - Data Model. In ISO/IEC JTC 1/SC34.*, January 2005 **http://www.isotopicmaps.org/sam/sam-model/**

[Kip03] Neill A. Kipp. *A mathematical formalism for the topic maps reference model. Draft paper submitted to ISO/IEC JTC1/SC34 Committee*, 2003, **http://www.isotopicmaps.org/tmrm/0441.htm**.

[Lea00] N. Leavitt. *Whatever happened to object-oriented databases?,*, 2000. IEEE Computer.

[LRH04a] Giovani Rubert Librelotto, José Carlos Ramalho, and Pedro Rangel Henriques. *Extração de Topic Maps no Oveia: Especificação e Processamento.*, In International Conference on Knowledge Engineering and Decision Support, pages 497-504. 2004, ISBN 9972-9876-2-0.

[LRH04b] Giovani Rubert Librelotto, José Carlos Ramalho, and Pedro Rangel Henriques. *XTche - A Topic Maps Schema and Constraint Language.*, In XML 2004 Conference and Exposition, Washington D.C., U.S.A, 2004. IDEAlliance.

[LRH06] Giovani Rubert Librelotto, José Carlos Ramalho, and Pedro Rangel Henriques. Metamorphosis - A Topic Maps Based Environment to Handle Heterogeneous Information Resources. In *Lecture Notes in Computer Science*, ISBN 3-540-32527-1, vol 3873, pages 14_25. Springer-Verlag GmbH, 2006.

[LSRH04] Giovani Rubert Librelotto, Weber Souza, José Carlos Ramalho, and Pedro Rangel Henriques. *Using the Ontology Paradigm to Integrate Information Systems.* , In International Conference on Knowledge Engineering and Decision Support, pages 497-504. Porto, Portugal, 2004.

[NLMa] U.S. National Library of Medicine. *MEDLINE - Fact Sheet.*, 2006, **http://www.nlm.nih.gov/pubs/factsheets/medline.html**

[NLMb] U.S. National Library of Medicine. *PubMed.* 2007 **http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed**

[NMB04] Mary Nishikawa, Graham Moore, and Dmitry Bogachev. *Topic Map Constraint Language (TMCL) Requirements and Use Cases. ISO/IEC JTC 1/SC34 N0548* 2004, **http://www.jtc1sc34.org/repository/0548.htm**.

[Omni02] Ontopia. *The Ontopia Omnigator*, 2002. http://www.ontopia.net/omnigator/.

[RLH05] José Carlos Ramalho, Giovani Rubert Librelotto, and Pedro Rangel Henriques. *Constraining Topic Maps: A TMCL declarative implementation.* In Extreme Markup Languages 2005, Montreal, Canada, August 2005. IDEAlliance

The Authors

Giovani Rubert Librelotto
UNIFRA
Santa Maria
Rio Grande do Sul
Brasil
giovani@unifra.br

Giovani Rubert Librelotto is a professor of Computer Science, Information Systems and Master in Nanoscience at Franciscan University Center - UNIFRA, Brazil. He received his PhD in Computer Science at the University of Minho in 2005, in Portugal. Prof. Librelotto has been involved in research around processing structured documents, XML, and topic maps. In the last years, he has been involved in several topic maps and bioinformatics projects.

Henrique Tamiosso Machado
UNIFRA
Santa Maria
Rio Grande do Sul
Brasil
htmachado@gmail.com

Henrique Tamiosso Machado holds a degree in Information Systems at UNIFRA, Brazil. He also is specialist in Administration of the Information Systems at UFLA, Brazil. Currently he is a Master in Nanoscience's degree candidate at UNIFRA, Brazil. He is Computer Science professor at URI, Brazil.

Mirkos Martins
UNIFRA
Santa Maria
Rio Grande do Sul
Brasil
mirkos@gmail.com

Mirkos Ortiz Martins holds a degree in Computer Science from UFSM, Santa Maria, Brasil and he is a Master's degree candidate from UNIFRA, Santa Maria, Brasil in Nanoscience. In 2003-2006 was teacher of Java Development from Web and Advanced Databases in IESVILLE - Technology and Cultural Institute, Joinville, Brasil. At same town, he was an Java Framework Analyst and XML researcher. Usually he is studing about quantum computer.

Pedro Gabriel Dias Ferreira
University of Minho, Computer Science Department
Braga
Portugal
4710-057
pedrogabriel@di.uminho.pt

Pedro Gabriel Ferreira's PhD research is focused on the analysis and extraction of sequence patterns also called motifs in protein sequences. These motifs occur in protein sequences because they have been preserved through the evolutionary history of the proteins. This is due to the fact that they most probably play a structural and a functional role in the protein's mechanisms. On the other hand amino acids outside these critical regions tend to be less conserved. The discovery of these motifs can be used to support a better understanding of the protein's structure and function. Besides, motifs are also important due to their wide-range of applications. Motifs can be used to perform clustering, family classification discovery of sub-families in large protein families, sequence annotation and the study and discovery of homology relations.

Currently he is working on the evaluation of significance measures for protein family classification and in the validation of sequence motifs through their tri-dimensional structure.

José Carlos Ramalho
University of Minho, Computer Science Department
Braga
Portugal
4710-057
jcr@di.uminho.pt

José Carlos Ramalho is a teacher at the Department of Informatics and a researcher at the CCTC research center.

He has a Masters on "Compiler Construction" and a Ph.D. on the subject "Structured Document Processing and Semantics". He is supervising several XML/SGML projects and acting as an external consultant for several institutions. He also has been the chair and chief editor of the portuguese XML conference.

Pedro Rangel Henriques
University of Minho, Computer Science Department
Braga
Portugal
4710-057
prh@di.uminho.pt

Pedro Henriques is an Associated Professor of Computer Science at University of Minho.

His research and teaching activity has been concerned with programming in general - paradigms, specification formalisms and languages; in particular, his main interest is the development of language processors.

He completed, some years ago, his Ph.D. at University of Minho in the area of Attribute Grammars; he is, now, the leader of the Language Specification and Processing group. The application of the grammatical approach to problem solving and the use of parsing and semantic analysis technologies in various problem domains (namely, document processing, information retrieval and data/text mining, and geographical information systems) are the present concerns of his academic work.

Extreme Markup Languages 2007®
Montréal, Québec, August 7-10, 2007
This paper was formatted from XML source via XSL
by Mulberry Technologies, Inc.