

TMS - TEI MANAGEMENT SYSTEMS

FEDERICO MESCHINI

*“That’s the trouble with computers, always think in black and white
No aquamarines, no blues, no imagination.”*
(Tom Baker, Doctor Who)

ABSTRACT

The Text Encoding Initiative Guidelines, used for the electronic encoding of literary and linguistic texts and since 2001 based on the XML syntax, are constantly being more and more adopted by the different digitalization projects. As the number of TEI digital texts and collections is growing, also the availability of software tools able to “manipulate” this kind of texts is increasing, and the choice is now much wider than it was just a few years ago. Each tool has its own peculiarities, which should be evaluated and confronted against the characteristics of the texts being encoded and the general needs and aims of the project of which the digital library is part. What this paper wants to analyze is the possibility of a classification of the actually available TEI Management Systems, using the Topic Maps technology, an ISO standard for the management and representation of knowledge.

1. INTRODUCTION

The expression *Digital Library* (DL) is really a fascinating one, and its charm is due both to its apparently oxymoronic nature and its actual meaning: the (re)creation in the world made of bytes of the place where human knowledge is preserved and made available. Moreover, such “intangible building” will take advantage of the features made possible by the digital dimension, from the multimedia features to the world-wide and 24/7 accessibility. Even the “life-cycle” and spreading of this term is very interesting: from being whispered during the seventies and the eighties only by a few researchers, academics and computer scientists, it is currently more and more known and used in different sectors of the information society, thanks also to the different librarian digitization projects undertaken and well publicized by the big players of the Net Economy such as Google, Yahoo and Microsoft [1].

Notwithstanding the commonly accepted definition formulated during an IBM workshop more than ten years ago [2], almost one hundred issues of a leading electronic magazine entirely on this topic [3] and several very relevant publications [4], when it comes to the concrete implementation of a *DL*, in my opinion the situation

resembles the old Indian story of the six blind men trying to describe an elephant [5], where each of the blind man gives his interpretation of the animal depending on the part of the body he has touched, and each time it is something different: a wall, a snake, a tree, a fan, a rope or a spear. In any case, this multi-semantic elephant, along with the chimera, a famous multi-composed mythological animal, could all be good metaphors to describe the over-layered composition of a *DL*. The factors in play are very different, both cultural and technical, and the latter ones are always changing. If we consider only the last few years of evolution of this sector, what results is a dynamic landscape: on one hand, the technical, where there is a constant renewal of the approaches and solutions, with a proliferation and affirmation of different standards, each centered on a particular aspect of this complex architecture; on the other hand, the cultural one, where the models of use and diffusion are extremely different, ranging from academic research projects, such as the *Women Writers Project* [6] or *Biblioteca Italiana* [7], to the institutional and disciplinary repositories conformable with the *Open Access Initiative (OAI)*, [8], passing by commercial projects, for example the *Safari Bookshelf* [9]. This feverish evolution does not seem to be slowing down, but rather the contrary and it is not easy to imagine what the actual situation in the medium and long-term will be.

2. XML AND (DIGITAL) LIBRARIES

Since the release of the first specification in 1998, the *eXtensible Markup Language (XML)*, [10]), has been adopted with notable success, even if with some hype, in several IT sectors. This has happened thanks to its potentialities, together with the family of related technologies, style-sheets and query languages [11], and all the currently available software, such as parsers, processors, frameworks and databases, by which it is possible to manipulate and work with the data expressed when using this syntax.

Due to its characteristics, *XML* is suitable for information which has both a document-centric nature, a textual document, including a data-centric one, and a bibliographic record, and this matches perfectly with the intrinsic needs of the *Library and Information Science (LIS)* sector. Previously in *LIS*, mark up languages were less important than other technologies, such as relational databases, but now the situation is changing very quickly. In fact by using *XML*, or better, when using the languages and standards built following its rules, traditional services are being rewritten in order to offer both traditional and new functions [12], and new services are arising, such as the creation and publishing of digital collections [13]. Depending on the kind of data, or better, on the nature of information, there is a particular *XML*

language, or “dialect”, to be used. Each one of these languages has its own design, but they all share a common underlying syntax, and many of the advantages of the use of *XML* derive from this factor. Hence, by restricting the list to the *LIS* sector, we have one or more “dialects” for textual encoding [14], metadata [15] and structural metadata frameworks [16], bibliographic records [17], confederate researches [18], archival finding aids [19], editorial information [20] and the creation of semantic networks [21].

With this young but already broad landscape, the use, or better, the affluence of *XML* in *LIS*, cannot any longer be considered a momentary trend, but has to be accepted as a natural evolution, related in many ways with the general redefinition of the sector, due to the increasing role that electronic resources are currently assuming.

3. DIGITAL (MEDIA) LIBRARY

Traditionally, the term “library” has always been related with information of textual nature, but in recent years the situation has greatly changed and nowadays it is perhaps more appropriate to speak of media library, as the other kinds of information, iconographic, audio and video, are also being preserved and made available. If this is true for “brick and mortar” libraries then it is even more for DLs, where there is only one base format: the long sequences of bits, composed entirely by 0 and 1, which are the common stuff every digital item is made of. Collections in a DL can contain both multi-format and multimedia items. Multimedia because the information being represented can be of any kind, on both a syntagmatic and paradigmatic level, and multi-format because, notwithstanding the common base of bits and bytes, the electronic representation is stored in one of the many possible formats. For example, a sequence of characters, which transmit textual information, can be memorized in one of the following forms: ASCII plain text, RTF, Microsoft Word, LaTeX or Adobe PDF. Moreover, each of these formats very often depends on a particular hardware and software combination, and this can be a limit for both preservation and dissemination issues. Solutions to this problem are the use of format approved by standard organizations, which are often hardware and software independent or, more realistically, the use of “de facto standards”[22]

As a logical consequence, the software used for the implementation of a DL must be able to manage the most possible number of formats. The result is a general “top-down” approach, where a single framework provides all the basic functions: the organization and browsing of the collections, usually organized in a hierarchical structure; the administration of the users, and the relative authorizations; full-text

and metadata searches and the retrieval of the selected items [23]. The format of the single digital object is one of the many allowed by the framework, which does little or even no operation on it, apart from memorizing its position: at a “physical” level in which the path of the file system, or table of the database, is memorized according to the storage method used; at a “logical” level, in which the collection of the DL is present and needs to be displayed to the users.

This approach is undoubtedly very practical, above all when you have to manage large quantities of items which are very different among themselves [24]. This approach, of course, is not the only one possible, and in many other situations perhaps not the most advisable, especially when dealing with a limited and particular subset, both concerning the kind and the domain of the information, as for example dramatic works written in England during the sixteenth century. In this case, another type of approach should be taken into consideration, with its own strong points and weaknesses, a “bottom-up” one, where the starting point is not the general framework but the single document, upon which all the architecture is built. This also involves major attention towards the preservation aspect, which usually becomes a main issue; for this reason, when dealing with textual materials of some cultural or scientific relevance, the markup of the actual content in XML is becoming a common and recommended practice in several projects.

4. TEXT ENCODING INITIATIVE

Nearing its twentieth anniversary, since it was initially launched as a research project in 1987, the *Text Encoding Initiative* (TEI, [25]) is currently an international consortium hosted by academic institutions and organizations. TEI’s main activity is the production, development and maintenance of the *TEI’s Guidelines* [26], for the representation in digital format, through text encoding and markup [27], of humanistic, literary and linguistic texts. Largely used and well established in different projects [28] *TEI Guidelines* have undergone a series of proposals, which give the name to the several releases, the latest of which has been called P5 [29]. Previously based on the *Standard Generalized Markup Language* (SGML, [30]), the *Guidelines* have used XML since the P4 version [31], released in 2002, and, given the complexity of literary texts, they currently adopt a modular approach, so as to organize the numerous elements and attributes used to encode the different textual aspects, whether related to prose, poetry, drama or linguistic and structural features. A simplified set of the *Guidelines*, called TEI Lite [32], is also provided, both for an introduction and for projects where a deep and complete markup is not required. TEI also plays a part in a very active and lively community [33], with a strong sense

of cooperation, and in several work [34] and special interest groups [35] regarding related topics, such as overlapping markup, ontologies, manuscript encoding and authors and presentation tools.

Presentation tools is a simple label for what should be defined as *TEI Management Systems (TMS)*: specialized *CMS* for *TEI* digital collections. In fact, if the Guidelines are very detailed and exhaustive, after the encoding everything else is left to the decision of the single project. During recent years many different solutions have emerged, each one with its own particular features and architecture. This has happened also thanks to the proliferation of software which is “XML-enabled”, and therefore can be adapted, with some customizations, to manage *TEI* documents, which are completely valid *XML* files. Therefore, for this purpose several kinds of software have been used: publishing framework [36]· *XML* native database [37]· more traditional *RDBMS* together with scripting or programming language [38]· search engines [39]· commercial frameworks [40]. As a natural evolution, and in answer to constantly increasing appeals, some frameworks have appeared lately which are specially designed for working with *TEI* documents [41], often based on some of the programs quoted above, according to the notorious principle of not “reinventing the wheel”, but only customizing it for a particular route.

Even if they all share a common “data layer”, composed by one or more *TEI* texts, each *TMS* has its own particular features and functions, and this is due both to the kind of the technology utilized and the nature of the project where the *TMS* was developed and implemented. Some tools offer multiple output feature and text/image editions, whereas others provide also forms of textual analysis, other than full-text and metadata search. Moreover, even the textual nature has some influence; in fact, a Medieval text is fundamentally different, in content and form, from a contemporary one, and has, therefore, different requirements for the manipulation of its electronic representation.

For all these reasons, there is indeed a strong need for a classification of the currently available *TMS*. But what should the principles of this classification be, and above all, how can they be achieved? Moreover, to be really useful this classification should cover all the many aspects which constitute a *TMS*, from the most technical, such as the hardware requirements or software specifications or the kind of programming language used, to the user-level ones, about the functionalities offered to the end user.

For all these issues, instead of using traditional technologies and methods, as the relational model used in databases, it could be interesting to try to explore new forms of representation and management of information and knowledge, in particular those which go under the label of “Semantic Web”.

5. SEMANTIC WEB AND TOPIC MAPS

Tim Berners-Lee, the main ideologue of the Semantic Web (SW, [42]), defines it as “[...] an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” [43], then making the resources available on the web machine-understandable, with an incredible range of new possibilities and services [44]. The W3C has created the key technologies of RDF [45], RDFSchema, OWL and SKOS to express the metadata, ontologies and semantic networks which are essential for the SW, but an interesting alternative is the Topic Maps (TM, [46]) model. TM are an ISO standard (ISO/IEC 13250:2003) for the knowledge management and representation, and described as “the GPS of the information universe” [47]. TM have their roots in indexes, cross-references, glossaries and thesauri in traditional media and are based on the simple but powerful Topics - Associations - Occurrences (TAO) paradigm. Topics represents the general and intangible subjects of which information is made (such as “William Shakespeare” or “Hamlet”), Associations are the relationships between Topics (“William Shakespeare is the writer of Hamlet”) which play a specific association role (respectively “is the writer of” and “is written by”). Finally Occurrences are the embodiment of Topics (such as a digital image representing William Shakespeare or an electronic or printed edition of Hamlet). Different TM about different topics can be merged together allowing a multi-dimensional approach to knowledge management. Thanks to their flexibility TM can be used in different contexts; currently the principal use is the creation of layered and complex navigation routes concerning existing resources thus giving them a new structure and information architecture, with conspicuous advantages [48]. On the implementation side there are currently two main syntaxes, XTM [49] and LTM [50], which are used to express a TM, and a growing number of related technologies for manipulation at a higher level, such as visualization and querying [51].

In TEI community TM seems to be the favourite SW technology, more than RDF/OWL. One possible answer is that when marking up a text following the TEI Guidelines, in some way you are working with implicit intra-textual relationships. Consequently, as a side effect, you start seeing extra-textual relationships as well and this is mainly what TM are about.

6. TOPIC MAPS FOR TEI MANAGEMENT SYSTEMS (TM4TMS)

For all these reasons a Topic Map could be an interesting way to describe the TMS, which are characterized by a complex and layered situation, where there are so many different aspects to be considered: technologies used, people involved, services offered

and so on. The first step is the creation of the conceptual model which describes the information domain that the TM should use as its scheme. This can be achieved in several ways. One possible way, and the one that I have followed, is the analysis of the actual landscape, using an inductive method, with the consequent creation of an Entity-Relationship diagram, which identifies all the topics involved and the roles played. To give an idea of the complexity of the situation, what follows are the descriptions of some of the “real cases” observed and analyzed:

1. the TEI texts are published in HTML, PDF and eBook format through an open source publishing framework which supports dynamic XSLT transformation, using some customized stylesheets. The textual queries are made possible through an open source search engine which is integrated with the publishing framework [52].
2. An open source Native XML Database is used for storing the texts. This database supports XQuery, then XQuery expressions are used for browsing and searching the text collections. The selected texts are published through XSLT transformation, using an XML publishing framework integrated with the database [53].
3. An XML-aware commercial Search Engine is used for storing the TEI texts, and all the other functions (publishing, searching and so on) are implemented using customized Perl scripts, instead of the one bundled with the software. For publishing the texts in HTML an XSLT engine written in Perl is used [54].

The next step was the creation of a general model based on information encapsulated in the previous sentences. This model is based on the following assumptions:

- A digital library offers one or more services.
- A software tool has one or more functions, based on various technologies and often integrated with other programs.
- A software tool function has one or more options.
- A digital library service is made up of one or more functions and options of one or more software tools.

Technically (or logically) speaking it is an n-ary association, an association between three or more topics. From the Digital Library perspective the association is: “A service is offered by a digital library using one or more software tools”. On the other hand, from the Software Tool perspective the association is: “A function is implemented by software tools using one or more technologies.”

In the Service Association Type (“a Digital Library offers a service of”) there are three different roles, as shown in figure 1:

- Subject – the digital library, the active role.
- Object – the service offered, the passive role.
- Means – the software, the way through which the service is offered.

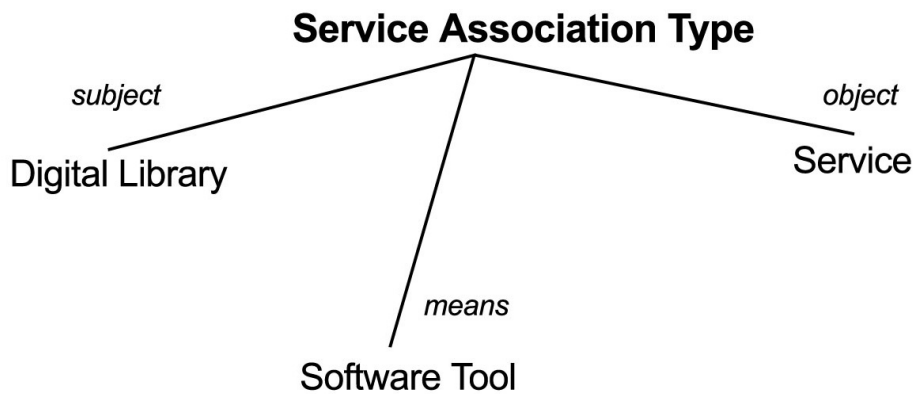


Figure 1: Service association type diagram

For a Function Association Type (“a Software Tool implements a function of”) there are always three roles, as shown in figure 2:

- Subject – the software tool, the active role.
- Object – the function implemented, the passive role.
- Means – the technology, the way through which the function is implemented.

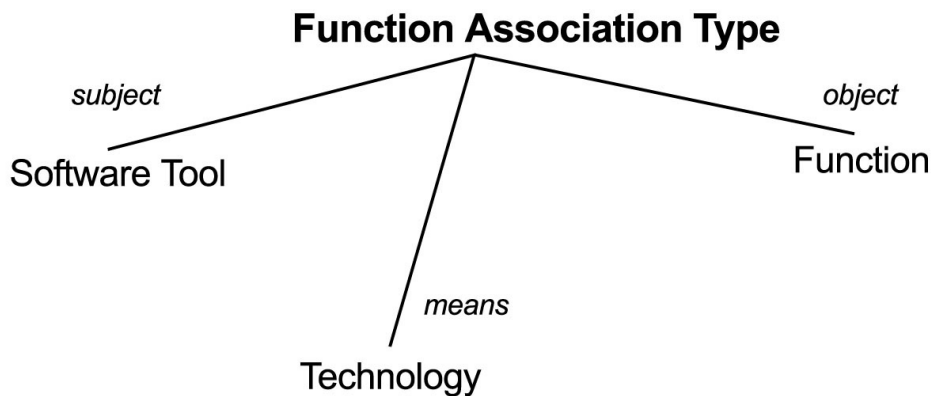


Figure 2: Function association type diagram

The association between Services and Functions is a binary one, where a service is “composed by” one or more functions, and one or more functions “compose” a service.

The four main [55] functions offered by TMS are:

- Publishing
- Search
- Text Analysis
- Administration

Each function has subclasses, or specialized functions, and options. For example, a full-text and metadata search are both subclasses of the general search function and some of the possible options can be used as support for boolean operators, result highlighting, delimited search, phrase search, proximity, research scope, wildcard, regular expression, stop words, and stemming. Regarding publishing, this function is the superclass of HTML and PDF publishing, based on the respective technologies. Some of the options for HTML publishing are: full publishing, when the original text is published in a single HTML page, regardless of its original length; chunked publishing, in this case the original text is divided into more than one file, according to its structure; text/image edition, when together with the electronic text there are also the images of the paper edition.

What has followed was the creation of a TM, using the XTM syntax, which captured the different aspects of the TMS in an essential and pragmatical way, what in other words would be called their ontology. Therefore, the Topics represent the computer standards, technologies and programs used, the digital libraries, the textual collections, the people and organization involved, all connected by several types of Associations, such as utilize, service, function, integration, supertype-subtype [56], requirement and support [57].

If we try to describe a concrete example we can take into consideration the case of Apache Cocoon, an XML framework [58], used for web publishing and development. In the following passage, near the subjects, objects and predicates there is a (T) or an (A) to mark their role as Topic or Association [59]: “Apache Cocoon (T), on a logical level, is a subtype of (A) a Framework (T), and, on a technical level [60], is a subtype of (A) a Servlet (T). A Servlet (T) is a subtype of (A) a Java Program (T) [61], and requires a (A) Servlet Container (T). Apache Tomcat (T) is a subtype of (A) a Servlet Container (T) [62]. Apache Cocoon (T) has integrated (A) the XSLT processors Xalan J [63] (T) and the XSL-FO processor FOP [64] (T) in order to implement (A) dynamic XSL transformation [65] (T) and has also integrated (A) Lucene (T) to implement (A) search functions (T). The New Zealand Electronic Text Center (T) utilizes (A) Apache Cocoon (T) together with (A) their customized XSLT stylesheets (T) to implement (A) the HTML and PDF publishing function (T) on their TEI texts (T) with the chunked publishing (T), full publishing (T) and text/image edition (T) options. Moreover Apache Cocoon (T) is integrated in (A) eXist [66] (T) in order to implement (A) the publishing function (T).”

This short example is just the top of the iceberg, but is enough to give an idea both of the complexity of the situation and the potentialities of a Topic Map to manage this complexity. During the presentation of this paper at ODOK '05 a demonstration of the TMS Topic Map was made by using the Omnigator, a free J2EE semantic tool created by Ontopia [67], which allows the browsing of both XTM and RDF file. Since version eight, Omnigator has also included the Vizigate feature, which allows a graphical representation of the Topic Map using the TouchGraph technology [68]. In figure 3 you can see the Omnigator on the page of the Apache Cocoon Topic, showing all related information and related association types, while in figure 4 the same information is visualized using the Vizigate.

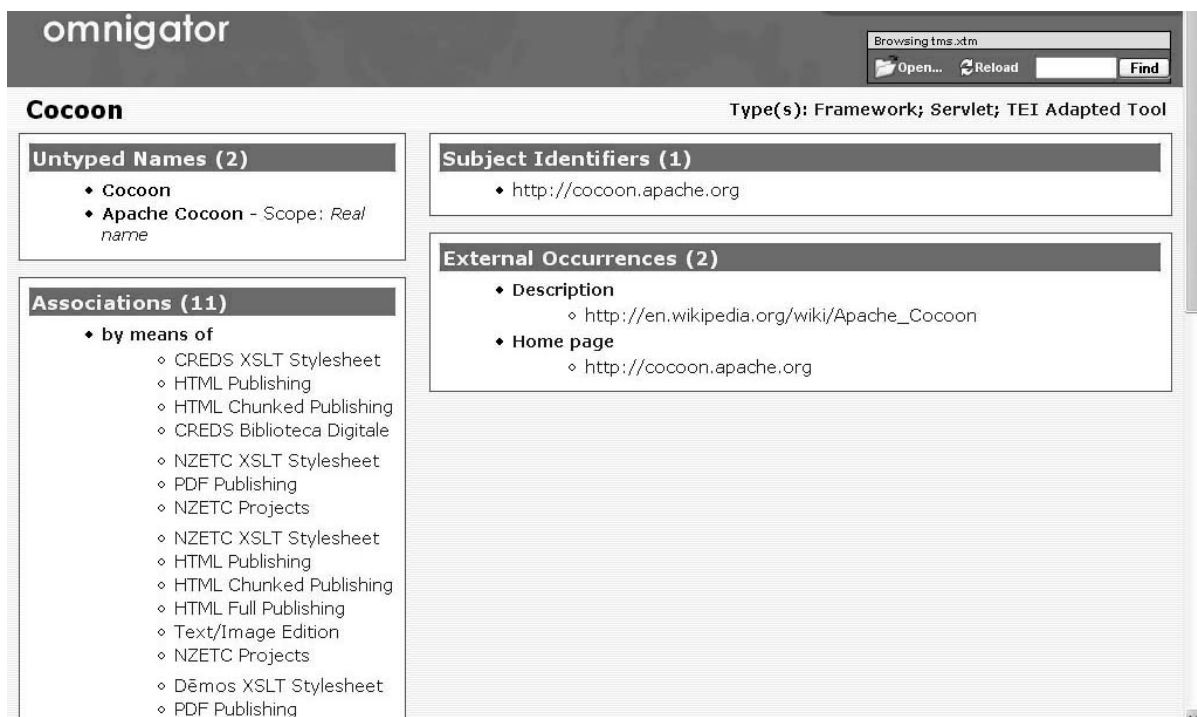


Figure 3: The Apache Cocoon Topic Page on Omnigator

7. CONCLUSION

The creation of a Topic Map concerning the Tools used on digital texts, encoded following the TEI's Guidelines, has been an extremely fascinating "case study" and it helped in underlying many issues which would have probably not emerge when using other methods. Naturally to make this model really helpful for the TEI community (but not only) it should be made available online with a more traditional web interface, and this idea has been already discussed in the Presentation Tools SIG.

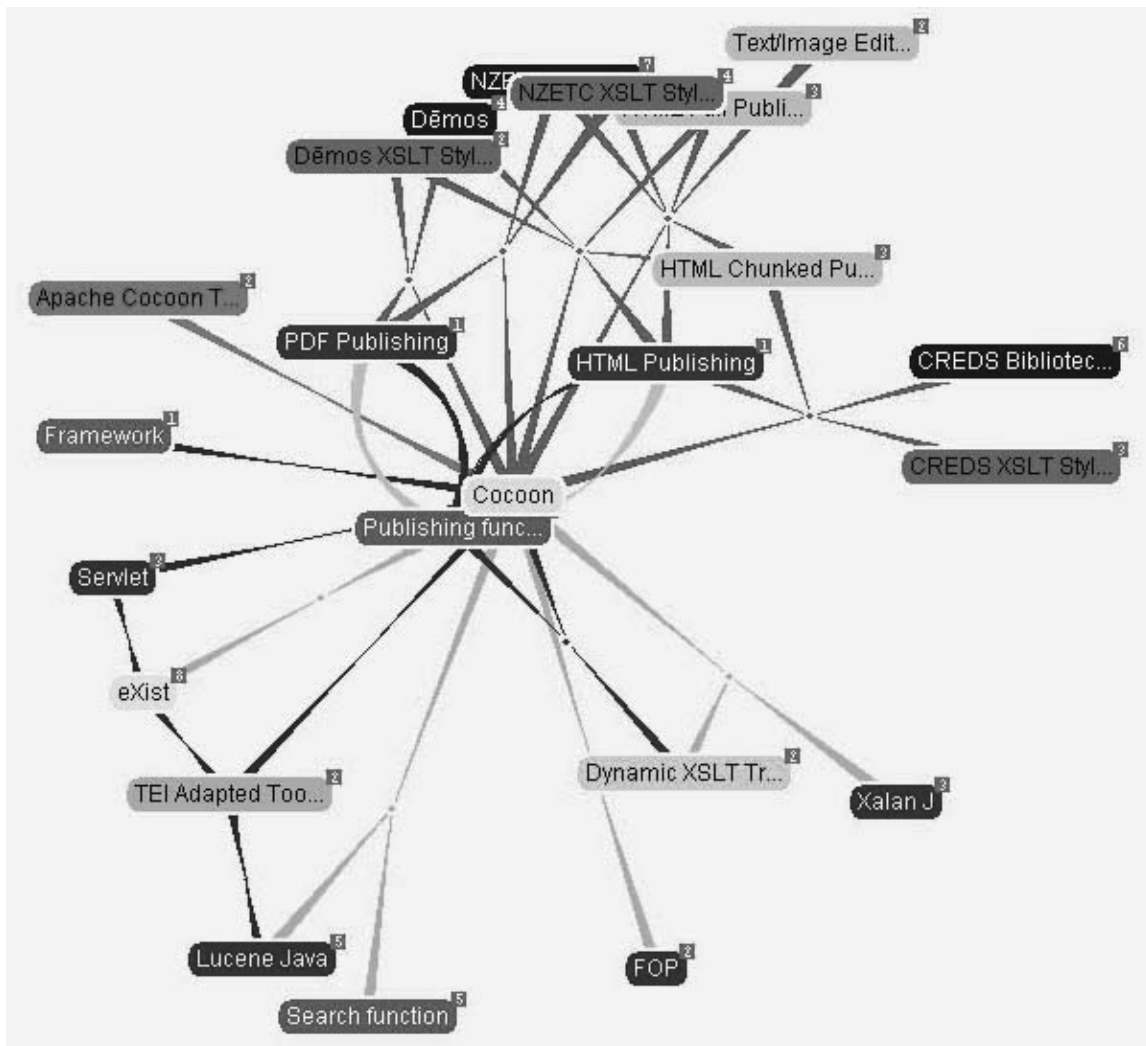


Figure 4: Vizigate visualization of Apache Cocoon Topic Page

In conclusion, in my opinion semantic technologies such as TM would be very relevant even in the medium term, and it would be very interesting to try to use them now with some other types of recent applications, such as AJAX [69], thus mixing together at the same time the both different and similar Web 2.0 and Semantic Web paradigms.

REMARKS

Web sites last access: 15/02/2006.

- 1 For an overview of the current situation see G. RONCAGLIA, I progetti internazionali di digitalizzazione bibliotecaria: un panorama in evoluzione, Res Cogitans, 2005, <<http://www.rescogitans.it/main.php?articleid=90>>.
- 2 "A digital library is a machine readable representation of materials which might be found in a university library together with organizing information intended to help users find specific information. A digital library service is an assemblage of digital computing, storage, and communications machinery together with the software needed to reproduce, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, storing, cataloging, finding, and disseminating information. A full service digital library must accomplish all essential services of traditional libraries and also exploit digital storage, searching, and communication". H. M. Gladney et al., Digital Library: Gross Structure and Requirements (Report from a Workshop), IBM Research Report RJ 9840, 1994, pg. 5, <<http://www.ifla.org/documents/libraries/net/rj9840.pdf>>.
- 3 D-Lib Magazine, <<http://www.dlib.org/>>.
- 4 Among the most recent, see M. LESK, Understanding Digital Libraries, Morgan Kaufmann (Elsevier), San Francisco - Amsterdam 2005 (2nd. ed.); L.A. TEDD and A. LARGE, Digital Libraries, Bowker-Saur, London 2004; R. TENNANT, Managing the Digital Library, Reed Press, 2004.
- 5 J. G. SAXE, The Blind Men and the Elephant, <http://www.noogenesis.com/pineapple/blind_men_elephant.html>.
- 6 <<http://www.wwp.brown.edu/>> .
- 7 <<http://www.bibliotecaitaliana.it/>>.
- 8 See <<http://www.openarchives.org/>>. For a comprehensive list of the currently active repositories see <<http://archives.eprints.org/>>.
- 9 <<http://safari.oreilly.com>>.
- 10 <<http://www.w3.org/XML/>>.
- 11 eXtensible Stylesheet Language Transformation (XSLT) <<http://www.w3.org/TR/xslt>> and XQuery <<http://www.w3.org/TR/xquery/>>.
- 12 As O.P.A.C., interlibrary loan, cataloging and indexing. At <<http://opac.bncf.firenze.sbn.it/opac/controller.jsp>> you can find the beta version of the new O.P.A.C. interface of Biblioteca Nazionale Centrale of Florence. This O.P.A.C. is built completely with open source software and the records are stored in a UNIMARCXML format. The schema is available at <<http://www.bncf.firenze.sbn.it/progetti/unimarc/slim/documentation/unimarcslim.html>>.
- 13 For more information see R. TENNANT (ed.), XML in Libraries, Neal-Schuman Publishers, New York, 2002, <<http://www.neal-schuman.com/db/0/290.html>>.

- 14 Overall the TEI < <http://www.tei-c.org/>> standard for literary texts and DocBook < <http://www.docbook.org/>> for scientific and technical texts.
- 15 DublinCore <<http://dublincore.org/>>.
- 16 METS < <http://www.loc.gov/standards/mets/>> and MPEG21 <<http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>>.
- 17 MARCXML <<http://www.loc.gov/standards/marcxml/>>, MODS <<http://www.loc.gov/standards/mods/>> and BOOKMARC <<http://www.bookmarc.pt/unimarc/>>.
- 18 OAI-PMH < <http://www.openarchives.org/OAI/openarchivesprotocol.html>>.
- 19 EAD <<http://www.loc.gov/ead/>>.
- 20 ONIX <<http://www.editeur.org/onix.html>>.
- 21 RDF<<http://www.w3.org/RDF/>>, RDFS <<http://www.w3.org/TR/rdf-schema/>>, OWL <<http://www.w3.org/TR/owl-features/>>, SKOS <<http://www.w3.org/2004/02/skos/>> and Topic Map <<http://www.topicmaps.org/>>.
- 22 A format which is largely used, so that the majority of the market considers it an authorized standard. Adobe PDF is considered a de facto standard.
- 23 There are currently many software forms available which provide such functions. Some of them are distributed with an open-source license and are also “Open Access compliant”. See Greenstone < <http://www.greenstone.org>>, DSpace < <http://www.dspace.org/>>, ePrints < <http://www.eprints.org/>>, Fedora <<http://www.fedora.info/>> and CDSware <<http://cdsware.cern.ch/>>. A list of open source and commercial software can be found at < <http://bcdlib.tc.ca/tools-software.html>>.
- 24 The Institutional Repository of M.I.T., currently contains approx. 20,000 items <<https://dspace.mit.edu>>.
- 25 <<http://www.tei-c.org/>>.
- 26 <<http://www.tei-c.org/Guidelines2/>>.
- 27 For more information on text encoding together with a description of TEI see A. RENEAR, Text Encoding in S. SCHREIBMAN, RAY SIEMENS, JOHN UNSWORTH (eds.), A Companion to Digital Humanities, Blackwell Publishers, 2005, pp. 218-239. For an example of TEI markup in action see <<http://computerphilologie.uni-muenchen.de/praxis/teiprax.html>> or <http://www.bibliotecaitaliana.it/archivio/leopardi/canti/leopardi_canti.xml>.
- 28 A list of the projects using TEI Guidelines can be found at <<http://www.tei-c.org/Applications/>>.
- 29 <<http://www.tei-c.org/P5/>>. P stand for “proposal”.
- 30 <<http://www.w3.org/MarkUp/SGML/>>.
- 31 SPERBERG-McQUEEN, C. M. and BURNARD, L., TEI P4: Guidelines for Electronic Text Encoding and Interchange. Text Encoding Initiative Consortium, 2002.
- 32 <<http://www.tei-c.org/Guidelines2/index.xml.ID=lite>>.
- 33 With a related mailing list: TEI public discussion list <<http://listserv.brown.edu/archives/cgi-bin/wa?A0=tei-1>>.

34 <<http://www.tei-c.org/Activities/index.xml.ID=wg>>.

35 <<http://www.tei-c.org/Activities/SIG/>>.

36 Cocoon <<http://cocoon.apache.org/>>, AxKit <<http://axkit.org/>>, Zope <<http://www.zope.org/>>.

37 eXist <exist.sourceforge.net/>, Xindice <<http://xml.apache.org/xindice/>>.

38 As the well-know couple MySQL/PHP, used on the larger part of web applications.

39 Lucene <<http://lucene.apache.org>>, XAIRA <<http://www.oucs.ox.ac.uk/rts/xaira/>>, XPAT <<http://www.dlxs.org/products/xpat.html>>.

40 DLXS <<http://www.dlxs.org/>>.

41 teiPublisher <<http://teipublisher.sourceforge.net>>, XPhilologic <<http://barkov.uchicago.edu/xphilo/>>, Anastasia <<http://anastasia.sourceforge.net>>, XMod <<http://www.cch.kcl.ac.uk/xmod/>>, eXtensible Text Framework (XTF) <<http://www.cdlib.org/inside/projects/xtf/>>.

42 <<http://www.w3.org/2001/sw/>>.

43 T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American, May 2001 <<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>>.

44 For more information see T. BERNERS-LEE, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web, Collins, 2000; G. ANTONIOU, F. VAN HARMELEN, A Semantic Web Primer, MIT Press, 2004; T. B. PASSIN, Explorer's Guide to the Semantic Web, Manning Publications, 2004.

45 For an introduction to the logical model of RDF see RDF Primer <<http://www.w3.org/TR/REC-rdf-syntax/>>.

46 For more information see: L. M. GARSHOL, What Are Topic Maps, <<http://www.xml.com/pub/a/2002/09/11/topicmaps.html>>; <<http://www.topicmaps.org/>>, the website of the Topicmaps.Org Consortium; <<http://topicmap.com/>>, a comprehensive website with news and information related to TM; J. PARK, S. HUNTING (ed.), XML TOPIC MAPS – Creating and Using Topic Maps for the Web, Addison-Wesley, 2002.

47 From the classical introduction S. PEPPER, The TAO of Topic Maps, finding the way in the age of infoglut, <<http://www.gca.org/papers/xmleurope2000/papers/s11-01.html>>.

48 A good example is the New Zealand Electronic Text Center <<http://www.nzetc.org>>, which has completely redesigned its structure and its digital collections using a Topic Map, based on the CIDOC-CRM conceptual model for the cultural heritage <<http://cidoc.ics.forth.gr/>>.

49 <<http://www.topicmaps.org/xtm/index.html>>.

50 <<http://www.ontopia.net/download/ltm.html>>.

51 The Ontopia Knowledge Suite <<http://www.ontopia.net/solutions/products.html>>; TMAPi <<http://www.tmap.org/>> a set of common API; TM4J <<http://tm4j.org/>>, an open source toolkit written in Java; TMQL <<http://www.isotopicmaps.org/tmql/>>, a query language and TMCL <<http://www.isotopicmaps.org/tmcl/>>, a constraint language;

- XTM4XMLDB <<http://sourceforge.net/projects/xtm4xmldb>>, an implementation of TMAPI for native XML database.
- 52 New Zealand Electronic Text Center <<http://www.nzetc.org>>.
- 53 CREDS <<http://matrix.scu.uniroma1.it/creds/biblioteca.htm>>.
- 54 The Swinburne Archive <<http://www.lettrs.indiana.edu/swinburne>>.
- 55 Using the terminology of Object Oriented Programming we could talk about “Super Class”.
- 56 The supertype-subtype concept can be considered equivalent to that of class-subclass.
- 57 Usually the topics are the nouns, while the associations are the verbs, just as in the Entities/Relationships diagram.
- 58 For the definition of a framework see <<http://en.wikipedia.org/wiki/Framework>>.
- 59 Following the T-A-O paradigm the Associations are bidirectional and they are considered Topics as well.
- 60 Technical level and logical level could be considered as the Scope of the Association.
- 61 And a Java Program (T) is a subtype of (A) a computer program (T) written using (A) the Java language (T).
- 62 An immediate consequence is that Apache Cocoon can be used together with Apache Tomcat.
- 63 <<http://xml.apache.org/xalan-j/>>.
- 64 <<http://xmlgraphics.apache.org/fop/>>
- 65 Dynamic XSL transformation (T) is a subtype of (A) XSL transformation (T), which requires (A) an XSL stylesheet (T) and a XSL processor (T).
- 66 eXist (T) is a subtype of (A) a Servlet (T) and a Native XML Database (T), which, in its turn, is a subtype of (A) a Database (T). Also a Relational Database (T) is a subtype of (A) a Database (T).
- 67 <<http://www.ontopia.net/download/index.html>>. An online demo with several Topic Maps is available at <<http://www.ontopia.net/omnigator/models/index.jsp>>.
- 68 <<http://www.touchgraph.com/>>.
- 69 Asynchronous JavaScript And XML. See <<http://en.wikipedia.org/wiki/AJAX>>.

ADDRESS OF THE AUTHOR

Federico Meschini

Università degli Studi della Tuscia

Facoltà di Lingue e Letterature Straniere Moderne

Viterbo, 01100 (VT), Largo dell'Università

E-Mail: fmeschini@tin.it