Liliana Maria Melgar Estrada

TOPIC MAPS AND LIBRARY AND INFORMATION SCIENCE:

An exploratory study of Topic Maps principles from a Knowledge and Information Organization perspective

ABSTRACT

Purpose: This master thesis attempts to present *a* 'state of the art' of the placement of Topic Maps (ISO13250) in Library and Information Science, through an extensive literature review and a synthesis based on their principles. It was sited from a Knowledge and Information Organization perspective, represented by the work by Elain Svenonius *The Intellectual Foundation of Information Organization* and some of the concepts of Knowledge Organization. This thesis also intends to present a conceptual and theoretical framework for future research.

Design/methodology/approach: The study under review presents a qualitative approach based on Grounded Theory principles to analyse the literature and build the conceptual framework for its analysis. The literature reviewed consisted of more than sixty documents, which included, among others, journal articles, conference presentations and papers, student reports and thesis, as well as a book chapter. Moreover, this was complemented with information obtained from mailing lists, blog postings and websites, and some unstructured interviews.

Findings: Topic Maps appears to be a development aligned within the tradition of Knowledge and Information Organization but is completely adapted to the context of the Web and the digital environments. In a LIS perspective, it is *bibliographic meta-language* able to represent, extend and mostly integrate all the existing Knowledge Organization Systems in a standards-based generic model applicable to digital content and online presentation.

Conceptually, Topic Maps is in the borders of the LIS discipline with Knowledge Representation and Computer Science, where LIS conceptual models play the role of intermediaries by providing the ontologies to the 'bibliographic universe'. Topic Maps questions traditional LIS views and principles. Even though some of them still remain the same, as the meaning-based identification of entities, the notions of 'document' and 'subject' require further studies.

Some important applications give account of the capabilities and potentials for further developments and research on Topic Maps in LIS. The main field of application is the Digital Humanities and TEI-codified texts presentation.

International Master in Digital Library Learning

Oslo, 2009

Acknowledgements

My sincere thanks to my dear supervisors.

I acknowledge Tóth Máté, Aki Kivelä, Sam Oh, Suellen Stringer-Hye, and Michelle Dalmau, and especially and Steve Pepper for their essential contributions.

Thousand thanks to my friends / colleagues at Oslo University College for the time shared. I am sincerely thankful to Elise for her help and valuable comments and to Bjørnar for the useful references and fruitful ideas. Also to Annette, Astrid and Tsigereda for their support and friendship.

I appreciate the service of the Oslo University College library that made possible this research.

I will always be thankful to the creators, coordinators, teachers, administrators and students of the DILL and Erasmus Mundus program.

Table of Contents

Tables

Figures

List of acronyms

AACR2: Angloamerican Cataloging Rules 2

AI: Artificial Intellingence

CC: Colon Classification

CDWA: Categories for the Description of Works of Art (Research at the Getty)

DCMES: Dublin Core Metadata Element Set

DDC: Dewey Decimal Classification

DOI: Digital Object Identifier

EAD: Encoded Archival Description

EAG: Electronic Archive Group

EATS: Entity Authority Tool Set

FRAD: Functional Requirements for Authority Data

FRSAR: Functional Requirements for Subject Authority Records

ICLO: Integrative Cross-Language Ontology

ILS: Integrated Library Systems

ISAAR(CPF): International standard archival authority record for corporate bodies, persons and
families

KO: Knowledge Organization

KOP: Knowledge Organization Processes

LCC: Library of Congress Classification

LCBS: London Classification of Business Studies

LCSH: Library of Congress Subject Headings

LCNAF: Library of Congress Name Authority File

LTM: Linear Topic Maps Notation

MARC: Machine Readable Cataloging

MADS: Metadata Authority Description Schema

MLA: Museums, Libraries and Archives

NKOS: Networked Knowledge Organization Systems

NZETC: New Zealand Electronic Text Centre

PMEST: Personality, matter, energy, space, time. The main facets in Ranganathan's Colon Classification.

PSD: Published Subject Descriptor

PSI: Published Subject Identifier

RDA: Resource Description and Access

RDF: Resource Description Framework

TIR: Thesaurus-based ontology information retrieval

TOIR: Topic Maps-based ontology information retrieval

TMDM: Topic Maps Data Model

UDC: Universal Decimal Classification

URI: Uniform Resource Identifier

VLE: Virtual Learning Environments (or Learning Management Systems)

W3C: The World Wide Web Consortium

OWL: Web Ontology Language

XSLT: The Extensible Stylesheet Language

XTM: XML Topic Maps

1. Introduction

1.1 Justification

Topic Maps[1] is an ISO standard for representing information about the structure of information resources (ISO13250). Its origins date back to 1991 when the Davenport Group started a project to develop DocBook, whose purpose was to facilitate the exchange of UNIX documentation using SGML/XML. A bi-product of this work was Topic Maps, a model and syntax whose original purpose was to enable the merging of back-of-book indexes.

The creator of Topic Maps, Dr. Steven R. Newcomb, had the insight that back-of-book indexes are actually a rendition of an underlying structure that could be represented explicitly. The initial model was consolidated and became an ISO standard in 2000. It immediately proved applicable to other domains, such as Information Architecture and Web publishing, and to enabling more structured (semantic) information retrieval on the Web. It is usually considered to be one of the technologies that facilitate semantic integration.

On the other hand, Library and Information Science (LIS) is a discipline with a long tradition and a body of knowledge quite structured and developed. Although the discipline is currently being challenged in its practices by new technologies, it is a time of flourishing for many of its traditional Information Organization principles and techniques, which are being applied in a variety of fields.

Topic Maps appears to be a development aligned within the tradition of these 'LIS techniques', but completely adapted to the context of the Web and the digital environments. However, its relatively new appearance and its origins in a different tradition from LIS (primarily the SGML community, which was concerned with document description languages), lead to a need for understanding its principles from an LIS perspective, both in order to explore how its concepts have been adopted and how its applications have been adapted to its practices.

[1] The convention that is followed in this text is, as in Pepper (2008b), to use "*Topic Maps"* (with initial capitals) to refer to the ISO standard and to its implicit model and conceptions; and "*topic maps*" to refer to the specific instances or artifacts created with Topic Maps.

The Topic Maps community has already taken some steps in this direction (e.g. Garshol, 2004), but from the LIS community, the efforts have been rather scarce or scattered. In this master's thesis, I present an overview of the integration issues and offer one possible approach to explaining Topic Maps from an LIS perspective, specifically from the Information and Knowledge Organization principles within the framework of what is known as "international librarianship" (Lor, P.J., 2008).

My interest in this subject matter originated during the course *Digital Documents* with emphasis on Topic Maps, which was given to the students of the International Master in Digital Library Learning at Oslo University College in 2007. Steve Pepper (one of the editors of the Topic Maps standard and one of its main leaders and disseminators) introduced the model and presented its main concepts related to the traditional bibliographic languages. At that time, the lack of examples of actual applications in the setting of libraries or other cultural heritage institutions and/or digital libraries or collections, in addition to uncertainties concerning the placement of Topic Maps within the LIS discipline, raised many questions that gave the motivation for the current work.

1.2 Background and significance

In 2002, Ahronheim (2002) said that the "work with topic maps is still too new to have produced papers describing library applications, but that will no doubt come in the next few years" (p.3). In 2005, Colmenero stated that the capabilities of Topic Maps haven't been fully explored because it is a relatively new standard[2] (p.98). At the time of writing this thesis, several publications and papers have indeed been produced, and some applications have been created. But it seems that nine years after the standard was published there is still no comprehensive research, or systematic study, of the various aspects of the integration of Topic Maps into the LIS discipline from a conceptual LIS perspective.

Most of the publications on Topic Maps in LIS present or report on specific applications, and those that approach the topic from a conceptual point of view do it mainly from a Knowledge Organization (KO) perspective. This topic appeared in the literature around 2000, the year in which the XML Europe conference in Paris brought together different communities interested in Topic Maps (Stringer-Hye, 2005; Sigel, 2003, p.452). A need for further research was also

[2] *"Por ser una norma relativamente joven no está aún totalmente explorado todo su potencial aplicativo"*

pointed out at that conference, to "interconnect Topic Maps with research directions in Information Science [or] Knowledge Organization" (Pepper, Sigel, Schmidt & Müller, 2000).

The most significant studies on the relation of Topic Maps in LIS from a Knowledge and Information perspective are, from the LIS community, Sigel (2003) and, from the Topic Maps community, Garshol (2004). Sigel (2003) made an extensive study on the relationship between Topic Maps and Knowledge Organization (KO). Exploring the principles of KO and Topic Maps, Sigel describes both sides of their relationship from a theoretical perspective (also adding some views on possible applications). Garshol (2004) made a comparative study of Topic Maps with the "common techniques from library science and information architecture" (p.16). He stated that Topic Maps is a relative newcomer to Library Science, but that it brings "the promise of better-organized web sites, compared to what is possible with existing techniques" (p.378). Garshol not only looked at Topic Maps in relation to work languages but also to document languages or *"metadata"*. The study was published in a highly reputed international Information Science journal and is commonly cited in the literature on Topic Maps in LIS. This master thesis is based to a great extent on these two papers.

In addition, Prof. Sam Oh, one of the most recognized researchers in the applications of Topic Maps to libraries, has made two presentations on the relationship of Topic Maps to the different models and schemas currently in use and under discussion in the LIS community: FRBR, RDA, SKOS and Dublin Core (Oh, 2008b; Oh, 2008c).

Iglesias & Stringer-Hye (2008) is the most recent study on the applications of Topic Maps to Integrated Library Systems (ILS). It presents "an overview of the current use of topic maps in the library field, how they might be integrated into the ILS structure and some of the inherent challenges in trying to transform MARC data […] as well as some ideas to where this technology will fit into the ILS" (p.12). Their research focused on collecting the practical applications of Topic Maps to library systems in The United States[3]. A limitation of this paper is the lack of explicit information about the research method, sampling, and theoretical frameworks used to draw the conclusions.

Chen & Chen (2001) report on a comparative study where Topic Maps was selected among classification and cataloguing systems, search engines, data mining, metadata and semantic

[3] (Suellen Stringer-Hye, personal communication, April 16, 2009)

networks to examine their origins, theories and methods, besides their current status, and trends. Unfortunately, this obviously significant paper couldn't be reviewed in detail by the author of this thesis because of language barriers.

Regarding the comparison of Topic Maps with other Knowledge Organization Systems (KOS), Kongsbakk (2004) made a detailed study of the similarities and differences between Topic Maps and thesauri, both from a theoretical and a practical perspective.

In terms of usability studies, Oh (2008a), Yi (2008) and Walsh & Dalmau (2006) observed that there are few usability studies on Topic Maps. Yi (2008) observed that, in general, few studies have examined user performance using ontology-based systems, and most of the existing research has employed RDF (p.1902). But some usability tests on Topic Maps appeared in the literature:

Yi (2008) conducted a usability test to explore how a Topic Maps-based ontology approach affects users' searching performance, trying to verify if it is positively affected by a system built with richer semantic relationships in XML and Topic Maps. For that purpose, he compared two systems: a Topic Maps-based ontology information retrieval (TOIR) system, and a thesaurus-based information retrieval (TIR) system in terms of recall and search time. Oh (2008a) conducted a similar study, comparing a Topic Maps-based system with an existing system for organizing and presenting information on Korean Folk Music. Walsh & Dalmau (2006), and Dalmau and Walsh (2007) report on a comparative usability study among Humanities researchers and students to see if "the inherent structure of a topic map-driven interface positively or negatively impacts the user's discovery process."Bøckman (2006; 2007) also reports on a usability study of a Topic Maps-based system prototype that was created to integrate the different vocabularies of specific domains in the Humanities. The overall results of the previous studies showed that the use of Topic Maps was positive

No comprehensive bibliography on the placement of Topic Maps in LIS was found either, with the exception of a project that has been discontinued (LITA interest group[4]). Other bibliographies, for instance in Topic Maps research[5], are broader.

[4] http://tm4lib.library.vanderbilt.edu/wiki/

[5] http://www.informatik.uni-leipzig.de/~maicher/bibliography.html

Many questions from an LIS perspective still remain unsolved in these studies: Is Topic Maps a KOS, or an information organization technique? If so, how does it relate to them, in which aspects does it differ, and which additional features does it provide? Does it represent an evolution of thesauri? Which is its way of representing knowledge? What is the application of Topic Maps to Knowledge and Information Organization in real settings, and how can it help improve information services?

No comprehensive research studying Topic Maps from an LIS perspective has been found that answers these questions, and the writings or experiences are until now scattered and lack a conceptual framework within which to conduct research.

1.3 Research design

The present work can be defined from its method as an exploratory and qualitative study on the placement of Topic Maps in LIS from a conceptual and practical perspective. It tries to answer a broad question on how Topic Maps integrates into the LIS concepts and applications. These research questions can be divided into the following sub-questions:

- What is the status of the incorporation of the Topic Maps standard in the concepts and practices of the LIS discipline?
- How can Topic Maps be explained from a conceptual LIS perspective?
- What capabilities of Topic Maps present possibilities for research and uses to address some of the current problems of organizing information in LIS?

1.3.1 Aims and objectives

The aims and objectives of this work are threefold:

a) To present a ‘state of the art’ of the placement of Topic Maps in Library and Information Science through an extensive literature review.
b) To understand the placement of Topic Maps in Library and Information Science from a conceptual point of view.
c) To identify some directions for further research and work on the incorporation and application of Topic Maps in LIS.

1.3.2 Methodological framework

This research is based on a qualitative approach within an interpretivistic epistemology. It was done in an iterative manner. Its purpose is exploration and description, but it does not pretend to be exhaustive. Some of the principles and methods of Grounded Theory (GT) research and analysis have been applied in this work.

1.3.3 Method

This study does not follow a formal qualitative method or strategy (such as case studies, action research or ethnography). It is basically an *'exploratory'* study[6] that uses some of the principles and techniques of the Grounded Theory (GT) approach.

GT is often referred to as a research method (Pickard, 2007, p.155) or as a methodology (Corbin & Strauss, 2008, p.1). It was originated in 1967 by Glaser and Strauss and has influenced qualitative research since then due to the suitable methodological implications of its epistemological and ontological assumptions for the kinds of realities and issues that are researched in the Social Sciences. Its nuances are varied and there are different implications for practical research. In this section, I only present what is essential to explain how this research was conducted.

Corbin & Strauss (2008) define the purpose of GT as "building theory from the data" and Pickard (2007) makes the distinction between GT as a method of qualitative research or as a qualitative data analysis technique. This is because, according to the GT epistemology, the researcher shouldn't approach the reality with a predefined coding or categorization of the object of research, but with the intention of letting these categories emerge from the data collected during the research, as if the theory was grounded in what is being researched.

Corbin & Straus assume the term in "a more generic sense to denote theoretical constructs derived from qualitative analysis of data" (p.1). In this master's thesis, GT was applied both as a data collection technique and as an approach to data analysis. It was in general, the approach followed in the research. As Pickard (2007) explains it:

[6] I assume here the term 'exploratory' in a non-technical sense, i.e. as the Oxford Dictionary defines the word 'exploration': "Travel through (an unfamiliar area) in order to learn about it". It is not used as it is in formal exploratory research, where "the term 'exploratory' refers to empirical work that has as a goal the discovery of new and unforeseen insight." (Seaman, 2007, p.23).

> "To take a grounded theory approach to research is to combine theoretical sampling, data collection, design of data collection, data analysis and theory generation, in one wholly interactive, iterative and interdependent process. The approach is an assemblage of all of these activities, which then allows a theory to emerge that is grounded in the data." (p.157, as cited in Pickard, 2002, p.17).

Since this approach was used to identify the placement of a concept (Topic Maps) in an existing theory (LIS), the role of the conceptual framework in the method was a key issue. According to Corbin & Straus (2008), the use of conceptual frameworks corresponds to a different methodology in research (the quantitative one) because they are previously developed and thus, contradictory with the idea of GT. These authors say that their preference is "not to begin our research with a predefined conceptual framework" (p.39). They consider, though, that a conceptual framework is useful in certain cases, among others:

- To frame the emergent concepts: "after studying a topic the researcher finds that a previously developed framework is closely aligned to what is being discovered in the researcher's present study, and therefore can use it to complement, extend, and verify the findings" (p.39);
- To offer alternative explanations.

This corresponds to the case in this master's thesis, where the first step was the selection of the sources, and in the middle of the process, the selection of a conceptual framework for the mentioned purposes. This was continuously and iteratively built until the end of the research, during the data analysis.

The role of the emergent categories (Appendix I) was to serve as indicators of a conceptual framework from LIS. The interpretation step consisted in making a synthesis of the emergent categories, building at the same time the conceptual framework based on them, which lead to the main conclusions. The main goal was to place a concept in existing 'theories' or conceptual frameworks that were discovered in the process as a result of the 'exploratory' approach already mentioned.

Those main conceptual frameworks corresponded to "Information Organization" and "Knowledge Organization (KO)" and many other related disciplines. For that purpose, I built an integrative conceptual framework to offer the alternative explanations. It is an attempt to map the concepts and terminologies of both disciplines and to integrate the other disciplines

and emergent concepts according to my interpretations. The result is graphically showed on page 39. This concept map was to helping me observing and concluding about the place of Topic Maps in LIS from a conceptual perspective, and to serve as a guidance for the reader of this thesis.

1.3.4 Data collection techniques

To collect the literature, I identified some basic sources, both from the Topic Maps community and from the LIS community, taking into account which ones were the most authoritative or widely used by the respective communities. These sources are summarized in Table 1. I considered "data collection" as the process of searching those sources.

For searching the sources selected in LIS, I used the words "Topic Maps", "topic map" and "ISO 13250". For searching the sources from the Topic Maps community, since the terminology within the LIS community is not consistent, I used some common expressions as "library techniques" and "information organization techniques", besides "library", "archive[s]", "museum[s]", "humanities" and "digital library" or "digital libraries". Moreover, I complemented the initial search with keyword searches using the terms from the conceptual framework to study the specific topics more in depth.

To select the literature from the initial search results I only chose the documents that had to do with LIS as a field and in the scope of MLA (Museums, Libraries and Archives) and when Topic Maps was applied to digital libraries. After refining the conceptual framework, I did a second selection based on it. That made me exclude documents that were, for example, in the field of Education, Information Architecture outside MLA, Knowledge Management, E-government, etc. Some of those initially retrieved documents were left in the selection only when they were relevant to support some important concepts in the synthesis, but they were not considered for the literature review.

I finally added to the literature, extra references that I found in the documents selected, or that I knew or were recommended to me, such as student papers and master thesis.

Table 1. Research Data Sources

Data Sources			Searching strategy (for TMs sources / For LIS sources)
Type	**From the TMs community**	**From the LIS community**	
Books and book chapters	Passin, T. B. (2004).	Cronin, B. (ed.) (2007)	Knowledge or Information Organization, LIS / Topic Maps
Journal articles (databases)	EBSCO (Academic Search Premier, LISTA, E-journals), EMERALD, WebOfScience	LISA	• TMs (subject and abstract fields) • ISO/IEC 13250 (subject and abstract fields) (from the results those articles related to LIS; libraries, museums and archives and their collections; digital libraries in the Humanities
Conferences (presentations and proceedings)	• TMRA • Topic Maps	• IFLA • ISKO • ALA • ECDL • IFLA • LITA	•Free search (Knowledge or Information Organization and specific concepts: faceted classification, etc.; libraries, archives, museums, digital libraries) / Topic Maps
Mailing lists	• TM mailing list • TMinLIS mailing list	• LITA (ALA) • ASIS • ISKO • NGC4LIB • DIGLIB	• Search in the archives (Knowledge or Information Organization and specific concepts: faceted classification, etc.; libraries, archives, museums, digital libraries) / Topic Maps • Request for information on applications
Websites	• Isotopicmaps.org • Topicmaps.org • Topicmaps.com • Coolheads.com • Techquila.com • Infoloom • Versavant • Ontopia • Ontopedia • Networked Planet • Topicmapslab.de	• OCLC • ALA • Dublin Core • IFLA • Digital Library Federation • DELOS • ISKO lifeboat	• Search in the archives (Knowledge or Information Organization and specific concepts: faceted classification, etc.; libraries, archives, museums, digital libraries) / Topic Maps
Blogs	• Lars Marius Garshol • Alexander Johannensen • Steve Pepper	• Alexander Sigel	Knowledge or Information Organization and specific concepts (faceted classification, etc.), LIS, "library organization techniques"
Bibliographies	• Bibliography of Topic Maps Research (2)	• Topic Maps for Libraries (1)	

(1) http://tm4lib.library.vanderbilt.edu/wiki/)
(2) http://www.informatik.uni-leipzig.de/~maicher/bibliography.html

I selected the applications of Topic Maps in LIS only in the context of MLA and digital libraries (as defined in 2.1.1 in the Humanities). For complementing the data collection I added to the literature search a request for information on applications to the mailing lists that appear in Table 1, and personal emails to some of the coordinators of these applications. Besides, I did three unstructured interviews to discuss some specific issues, with Prof. Sam Oh, Suellen Stringer-Hye, and Aki Kivelä.

From the sources I finally selected, I only reviewed those in English, Spanish, Norwegian, Danish, and Italian. For articles in other languages, I made the review based on the English abstract.

1.3.5 Data analysis

As already mentioned, I used GT as an approach to research, and additionally I used it as a way to analyze the data. I followed these steps as part of this iterative approach:

1. Identification of the sources
2. Data collection from data sources
3. General reading of all the documents compiled
4. First identification of possible conceptual frameworks
5. Codification of documents using a software for qualitative data analysis[7] (first topics emerged). Use of memos for annotations and analysis
6. Use of the software for creating categories (families of codes) and grouping the initial emergent concepts into groups (families). These were the emergent concepts (see Appendix 1 and section 5.1)
7. Selection of a conceptual framework in LIS according to the categories that emerged
8. First writing attempt and creation of the structure of the thesis according to the conceptual framework and the initial concept map
9. Second data collection (from references in those documents and based on the conceptual framework)
10. Second codification of documents
11. Refinement of the analysis and synthesis based on the conceptual framework selected in step 7 and the final conceptual map
12. Final refinements

[7] Atlas.ti (www.atlasti.com). See section 5.1 for a methodological observation.

1.4 Outline

This thesis is divided into six chapters. A concept map on page 39 serves as a guide to them and to their different sub-sections.

Chapter 1 has presented the introduction to the topic of research and the details on how it was conducted.

Chapter 2 presents a conceptual framework from the LIS discipline and related disciplines that was built during the research process according to the method indicated in 1.3. It serves to explain the place of Topic Maps in LIS.

Chapter 3 is an extensive literature review, a *state of the art* of the concerns and applications of Topic Maps in LIS. It is presented using the conceptual framework as an organizing principle.

Chapter 4 presents a synthesis of Topic Maps as seen by the LIS discipline, using the conceptual framework as a basis. In section 4.1 the main elements of 'bibliographic languages' serve as a point of departure to explore the main constructs and principles of Topic Maps. Section 4.2 and 4.3 conclude the exploration made in 4.1 presenting the place of Topic Maps in LIS in a conceptual perspective; and section 4.3 describes how Topic Maps can enhance Information and Knowledge Organization as a result of the analysis presented in the previous sections.

Chapter 5 discusses some of the issues that were also found to be the concern of the documents analyzed, but that were out of the scope of this thesis. It also presents a discussion of the methodology, results and implications of the present research.

Finally, **Chapter 6** summarizes the main conclusions of this work and presents the main possibilities for future research on Topic Maps in LIS. These directions are presented in a theoretical framework with some examples of possible research topics.

The **References** include both the literature on Topic Maps in LIS that was collected for the study, and the background literature. Both are presented together in alphabetical order, but the literature on Topic Maps in LIS that is reviewed in Chapter 3 is marked with an asterisk.

2. Conceptual framework

To understand the place of Topic Maps in LIS, I needed to choose a specific framework. This is because LIS is a multidisciplinary area of study with different research traditions and where diverse terminologies coexist. This framework had to be chosen according to the topics of concern found in the literature (as I describe them in 1.3.3). The conceptual frameworks that 'emerged' from the literature corresponded to Information Organization and/or to KO. This was expected since Topic Maps is defined as "a standard to represent information about the structure of information resources" (ISO/IEC 13250-1).

Information Organization derives from the tradition of Anglo-American descriptive and subject cataloging; while KO, comes from a long tradition derived from Information Science[8], a discipline with which Topic Maps share common theoretical principles (Colmenero, 2005, p.78). KO is the object of study of an international organization, ISKO, which was founded in 1989 (Dahlberg, 2007, p.84). In the LIS community, these two disciplines are usually referred as to 'Knowledge and Information Organization'.

Due to these different roots of Information Organization and KO in different researching traditions and to the changing nature of the LIS discipline (Zins, 2006) there is no agreement in the LIS community on their terminologies and some of their concepts. Besides, other disciplines as Information Management, Knowledge Management, and Information Retrieval, share some principles and concerns but not always the concepts and terms to refer to them.

For that reason, I needed to select a single approach to Information Organization and/or KO among the LIS discipline. I selected Elaine Svenonius' "Intellectual Foundations of Information Organization", which I found to be accepted and cited repeatedly by both the Topic Maps and the LIS communities. Her work is among the most comprehensive in the area[9], and is also closely related to the development of the FRBR model, one of the most influent and accepted conceptual models in the LIS discipline. I did not choose KO because it was not sufficiently specific with respect to descriptive cataloging, and because its concepts

[8] Information Science was known as Documentation until World War II (Day, 2008), and it's still called so in Spain.

[9] In the paperback edition of the book Barbara Tillett, one of the main figures in LIS nowadays writes: "This book provides sound guidance to future developers of search engines and retrieval systems. The work is original, building on the foundations of information science and librarianship of the past 150 years."

are scattered in different, equally important works which would add problematic conceptual issues to the selection of a framework. However I included some basic concepts from the KO discipline when Svenonius' concepts were not enough to explain Topic Maps in LIS. The main sources for these concepts from KO were Hjørland (2006a; 2008). Section 2.7 presents a summary of this synthesis, made for the purpose of using a coherent terminology along this text.

Following there is a description of the main concepts that are relevant to describe the position of Topic Maps in LIS. They constitute the conceptual framework of this thesis, which is presented graphically through a concept map in page 39. The map serves as a guide to navigate through these concepts.

2.1 Library and Information Science

Library and Information Science (LIS) covers two disciplines: Librarianship and Information Science. Librarianship focuses mainly on libraries as institutions and on the services to library users, while Information Science focuses, firstly, on the "nature of the information transfer process" and, secondly, on "information organization and dissemination through appropriate intellectual apparatus and technology" (Feather & Sturges, 2003). Nowadays, however, the term is used undifferentiatedly and is often used as a synonym of Information Science (Feather & Sturges, 2003, p.371). Although this equivalence is not a wholly accepted view, it is the one that will be assumed in this thesis.

In this perspective, LIS is an interdisciplinary field of study which relates, among others, with Information Management, Knowledge Management (understood as a branch of the business discipline), Information Retrieval, Information Architecture and Computer Science.

2.1.1 Digital Libraries

One concept that is commonly found in the literature is Digital Libraries. Although several definitions exist, they belong to different perspectives. For the purpose of this thesis, I will present my interpretation of the different meanings found in the literature. Those will be grouped into five more general perspectives:

1. The use of LIS techniques (Information Organization + KO) in the organization of digital content and Websites: "At its most basic, a digital library can be defined as

digital content that is organized along bibliographic principles" (Pepper, 2008b). This view corresponds to what defines, in one perspective, Information Architecture: "The art and science of structuring and classifying 'web sites' and intranets to help people find and manage information (Rosenfeld & Morville, 2002). However, not all Web sites are digital libraries.

2. The digital objects and the systems that organize them: refers to digital libraries as other types of information systems. These also include ILS in the digital environment. This view is described by Aalberg (2003) who considers that the surrogates of the physical or digital entities, what is called their metadata, are digital objects as well and thus (this is my assumption) constituents of a digital library:
3. "Other artifacts may be represented through other surrogates like a metadata record or merely an identifier. Such surrogates can still be useful in the initial steps of acquiring, discovering, evaluating and selecting information, and should be considered equal to digital content objects when discussing digital libraries" (p.9).
4. I consider that this view could be related to the 'digital library system' in the DELOS reference model (Candela et al., 2008).
5. The collections made available digitally: often digital libraries are used to refer to small or specific collections in a library or a bigger organization that are made digital. This perspective relates to the next:
6. The institutional repositories: digital libraries are often used as synonym with institutional repositories and the systems used to maintain them (DSpace, Fedora, etc.)
7. The organization (previously traditional libraries). Libraries as organizations have been qualified according to their technological development. Previously called "automated libraries" or "hybrid libraries", now they are referred as "digital libraries". This view is represented by the definition of digital libraries of the Digital Library Federation (DLF) and of "digital library" in the DELOS reference model (Candela et al., 2008).

I considered the second and third perspectives as the most relevant for this thesis.

Another concept of relevance in LIS and digital libraries, which is one of the main concerns found in the literature, is that of interoperability. Although this is an area of research in its own right with a large conceptualization, I have assumed some very basic definitions for the framework of this thesis. They are: 'metadata interoperability' (basically at the schema level),

as interoperability between document languages metadata schemas (e.g. metadata schemas); 'semantic interoperability', as interoperability between subject languages (e.g. thesauri, subject headings) and between subject and document languages; and 'linguistic interoperability', as the ability of systems to give access to multilingual applications.

2.2 Information Organization

2.2.1 Introduction

Elaine Svenonius (Svenonius, 2000[10]) presented "the systematized body of knowledge that constitutes the foundation of the systems designed for organizing information" (p.ix). For her, Information Organization is a body of knowledge with principles, objectives and techniques that has been developed within the LIS field, which includes nowadays the organization of information in digital form (p.xiv). Information Architecture could be an example.

Information Organization, in Svenonius' perspective, consists in the use of a particular "special-purpose" language to describe the information and its physical embodiments with the idea of accessing both of them. The languages used for that purpose are called "bibliographic languages", as opposed to "natural languages". As she sees it, Information Organization would consist of the "colossal labor of […] having to construct an unambiguous language of description –a language that imposes system and method on natural language and at the same time allows users to find what they want by names they know" (p.14).

In a later work (Svenonius, 2004), she changes the word 'bibliographic languages' to 'retrieval languages,' preserving basically the same meaning (not defined explicitly by her), which can be assumed to be an artificial language, a subset of natural language, designed for the specific purpose of embodying knowledge representations.

In the following subsections, the main concepts of Svenonius (2000) will be described as she presents them in her book, with some additional examples where needed.

[10] In 2.2.1, page numbers, if other references are not specified, refer to Svenonius (2000).

2.2.2 Bibliographic languages

The purpose of bibliographic languages is, in the Svenonius' view, to describe 'bibliographic entities', which are basically "works, editions, authors, and subjects" (p.31) divided in two realms. The first realm is that of 'information', the second one is that of the 'information entities' or the physical embodiments of the former. For Svenonius information is equivalent to "the content of a message" (2001, p.7), that is, the content of documents (which are defined as "physical embodiments" of information). "Document languages" is the term that she uses to name the bibliographic languages that describe such physical embodiments, and "work languages" those that are used to describe the information entities.

Most bibliographic languages are based on the concept of "description" of information and its physical embodiments. "A description is 'a statement of the properties of a thing or its relations to other things serving to identify it.'." (Svenonius, p.54). Bibliographic languages are mainly used as tools for making descriptions. Those descriptions are organized in bibliographic systems (indexes and catalogs).

In Svenonius (2004) the term "bibliographic languages" changed to "information retrieval languages".

2.2.2.1 Document languages

Document languages describe and identify manifestations, the attributes that are specific to particular physical objects that embody or manifest a work (Svenonius, p.137).

Although this definition was assumed originally in this thesis, some of the findings (chapter 4.2) had implications on it and showed that it is problematic: Svenonius' herself states, that objects of description can be treated either as attributes or as entities (Svenonius, 2000, p.53). For instance, an author could be represented either as an attribute or as an entity, i.e. when it appears as an element pointing to attributes of entities in a bibliographic description, or when it is treated as an entity in its own right. In the first case the author's name (or the property 'author') belongs to the 'document languages' realm, and in the second case is when an author name could be considered part of a 'work language'**.**

This would lead us observe that instead of thinking of document languages as instruments for describing embodiments of entities (that is, documents) they could be considered as such

when they describe any kind of entity. An author for example can be described using a MARC schema, in which case, that 'document language' is not being used to describe any physical embodiment. Section 4.2 goes into some details.

2.2.2.2 Work languages

Work languages describe the intellectual attributes, independent of any space-time manifestations, of information entities and provide access to information in documents (p.107). It can be characterized as an "information-content language" and as an "intellectual-attributes language" (Svenonius, p.53), this means that work languages describe 'information' which is contained in documents, specially the information that is used for their description, that is, to describe documents in terms of their authors, titles, editions and subjects.

Work languages include mainly **author** and **title languages**, **edition languages**, and **subject languages**. An author language is one that is meant to describe persons and corporate bodies associated with works (p.54), while a title language does the same for their titles (p.93) and an edition language for the editions. An edition is "each subclass of a work set formed by intellectual attributes." (p.97). Editions refer basically to "manifestations" in the FRBR model (see 2.4.2), and won't be studied in this thesis.

Among the work languages, **subject languages** are a special kind with respect to their degree of development and the extension of their use. They are "used to depict what a document is about" (p.127). They are of two types: "alphabetic languages" and "classification languages" (among the last ones, Svenonius includes as main examples of DDC and UDC). Alphabetic languages are "alphabetically-based" and classification languages are "classificatory or notationally based". They differ in (Svenonius, p.128):

- the way they designate subjects: verbal expressions vs. notations
- the use of different bibliographic systems: thesauri and subject authority lists vs. classification schemes
- the way to display and order the subjects: alphabetically vs. systematically (discipline plus hierarchical display of topics).

Library classification schemas (Batley, 2005) can be (although there is not a clear divisory line between them) analytico-synthetic (bottom-up classifications, like Colon Classification and Bliss Classification) or enumerative (top-down classifications like Dewey, LCC). UDC

shares both characteristics. Enumerative classifications, as the word describes them, try to enumerate, in a predefined way, all the possible subjects the materials are about, allowing the person who classifies to choose the appropriate subject (number). On the contrary, analytico-synthetic classification schemes or faceted classifications (as they are also known) do not try to predefine the notations; instead they provide basic concepts that can be combined or 'synthetized'. Its name comes from that process (the syntezis) and the previous analysis, which is used to decompose each subject into its basic concepts.

Classification schemas have then notations and subjects as vocabularies. These are structured, according to the different classification schemas, in schedules and subject indexes. The schedules are classified listings of subjects with their associated notations, while the subject indexes are presented in notation order. "Notation is the group of symbols, technically applied, which as a code represent the subjects contained in the schedules of a classification scheme in order that these subjects will be filed at the correct point in a physical sequence of subjects. (Batley, 2005, p.9). Notations can be pure (if it uses only one type of symbols) or mixed (when it uses a combination of them).

2.2.3 Components of bibliographic languages

The **components of a bibliographic language** (as they are also for natural languages) are its vocabulary, its semantics, its syntax and its pragmatics.

2.2.3.1 Vocabulary

The **vocabulary** of bibliographic languages consists of the simple and complex expressions used to name the values of the three variables: entities, attributes, and relationships" (see 4.4.). This means the terms or codes of the bibliographic languages that are available for use (Svenonius, p.55). An example could be the descriptors in a thesaurus (i.e. *audiovisual librarians* in the Art and Architecture Thesaurus), the terms for the classes in a classification schema (i.e. *791 Public performances* in the DDC), the title or author of a document in a bibliographic record or in an authority file (i.e. *Manovich, Lev* or *Lev Manovich*).

2.2.3.2 Semantics

The **semantics** in Svenonius terms refers to the "different meaning structures found in languages" (p.56). She identifies three of those structures:

- **Relational semantics** which refers to the meaning relationships among terms (for example, to the types of associations established in a thesaurus).
- **Referential semantics** which covers the "techniques used to limit the meanings or referents of terms" (p.57). In practice it applies, for example, to disambiguation rules.
- **Category semantics** which "has to do with the facets or grammatical categories into which the vocabulary is partitioned" (p.57). Those facets indicate that the terms that belong to them have the same or similar type of referents. As examples, Svenonius (p.57) cites the *topic*, *place*, *time* and *form* categories used by the LCSH.

2.2.3.3 Syntax

The **syntax** refers to "the ordering relationships among the component elements of complex expressions in the language" (p.55), to the "rules [to] specify the order in which individual vocabulary elements of the language are concatenated to form larger expressions." (p.58). In a bibliographic language "a well-formed string or heading [the equivalent of a grammatically correct sentence in natural language] is one that is authorized by syntax rule" (p.58). The syntax expresses "contingent relationships" (Svenonius, 2004), that is, the relationships that are context-dependent are specified using word composition rules that apply to the bibliographic languages (for example, the AACR2, or the ISBD rules). This is the difference, as explained by Svenonius (2004) between "syntagmatic or context-dependent" relationships as opposed to "paradigmatic or context-free" (i.e. definitional) relationships. Svenonius cites the example of the relation between *parrots* being *birds* (paradigmatic) opposed as *parrots* being *pets* (syntagmatic).

The syntax of a bibliographic language has to do with term-string composition, citation order, precoordination and postcoordination. An example of syntax in DDC, cited by Svenonius (2004), would be *323.4430976 [Free Speech in the South Central U. S.]*. Another example of syntactic composition could be seen in a Library of Congress Subject Heading: *Argentine literature--20th century—Film and video adaptations*.

2.2.3.4 Pragmatics

The **pragmatics** deals with the use or application of the language (p.58), that is, with the "rules for making descriptions", i.e. the cataloging or indexing rules that specify which elements should or not should be included in a description, when to create a new one, or how many of them including in a description (indexing depth). According to Svenonius bibliographic languages have a fairly underdeveloped pragmatics, except DDC which is highly populated with notes, explanatory definitions and instructions (p.58).

Since this component is not treated in depth in her book, I took the concept of Knowledge Organizing Processes (KOP) form the KO discipline. Those are, together with KOS the main objects of study of KO in a narrow perspective, as it is defined by Hjørland (2008). The main KOP in LIS are "abstracting, cataloging, classification, indexing, linking etc." (Hjørland, 2006). In this thesis I selected only three for the purpose of exemplification: cataloging, classification and indexing.

2.2.4 Bibliographic objectives

According to Svenonius, bibliographic systems organize the bibliographic descriptions which are made using bibliographic languages. Those descriptions are kept in bibliographic records. These systems are built based in different objectives: although she agrees that "no single [bibliographic system] need[s] aspire to meet all the needs of all users" (p.28), she also points out that there are common objectives that those systems need to prosecute and establish before they are designed. She draws attention to the background of the establishment of those objectives, going back to Panizzi, Cutter (who "in 1876 made the first explicit statement of the objectives of a bibliographic system"), and Lubetzky, who reviewed Cutter's objectives. His revision was adopted at the Paris Conference on Cataloging Principles held in 1961. They were reformulated in 1997 by the International Federation of Library Associations and Institutions (IFLA) which included them in the FRBR model as "User tasks". The user tasks are those generic tasks "performed by users when searching and making use of national bibliographies and library catalogues"[11]. They are: "to find entities that correspond to the user's search criteria […], to identify an entity […], to select an entity that is appropriate to the user's needs […], to acquire or obtain access to the entity described." (Svenonius, p.17).

[11] http://archive.ifla.org/VII/s13/frbr/frbr_current2.htm

Svenonius adopts the IFLA objectives but separates the "finding" objective into *finding* and *collocating*, and adds a fifth objective: *navigation*. For Svenonius, collocating is a key concept in the bibliographic discourse and one that can explain the overall purpose of Information Organization, that is, "to bring like things together" (p.18). Her five objectives of a "full-featured bibliographic system" –the system that fulfills all the objectives–, presented in the following sections, are still relevant in digital libraries and other types of information systems which contain more than bibliographic descriptions.

The term "bibliographic systems" is equivalent to the concept of 'KOS', both referring to what the Topic Maps community usually calls "library techniques", that is, according to Hjørland (2006a) the "tools that present the organized interpretation of knowledge structures."

2.2.4.1 Finding

Refers to the user need of finding a specific entity to which she/he knows the author, title or subject. "To *locate* entities in a file or database as the result of a search using attributes or relationships of the entities" (p.18). The traditional *finding objective* specifies that what is to be found is a set of documents, defined by the mentioned criteria (p.17).

2.2.4.2 Collocating

Cutter defined it in terms of the need for "sets" of documents on a given author, subject or genre (p.16). Lubetzky specified it to the "works" as basic units for the display. **Collocation** is defined by Svenonius as "the essential and defining objective of a system for organizing information". It consists on "bring essentially like information together and to differentiate what is not exactly alike" (p.11). This means that only records that are relevant are supposed to be brought together, not intermixed with those that are irrelevant. In the Cranfield experiments terminology, this is what is called "precision" (p.22). In general terms it consists in presenting together all documents that belong to a certain search criteria, especially to a given work, edition, author or subject.

The means through which collocation is achieved are related to the issues of identity. The term 'collocation objective' is extensively used in the Topic Maps literature.

2.2.4.3 Navigating

To find, in a bibliographic system, related works and attributes based on their associations. Svenonius uses Pierce Butler *navigation* methaphor of the *bibliographic universe*, which "is apt in its deiction of a user roaming from point A to point B and so on to reach a destination – the desired document–." (p.19).

2.2.4.4 Choice

It corresponds to the task of "identifying" an entity in the results, that is, to confirm that it corresponds to the entity sought or to distinguish between two or more entities with similar characteristics. "It assumes a user is faced with a number of similar documents and needs to make an effective choice from among them, such as from among several editions of a work" (p.16). It corresponds to the IFLA objective of "select[ing]" an entity that corresponds to the user's needs (content, physical format, etc.).

2.2.4.5 Acquisition

To obtain access to the entity described. There is a difference between intellectual access vs. physical access (Svenonius, p.122).

2.3 Knowledge Organization

2.3.1 Introduction

Sigel (2003) defines KO as "an interdisciplinary field that reflects the practical activity of organizing knowledge for specific purposes and discourse communities" (p.385). It is concerned with "organizing objects of thought (and associated carriers of information)" with the purpose of facilitating access through the use of "sophisticated finding aids". Tennis (2008) defines KO as "the field of scholarship concerned with the design, study, and critique of the processes of organizing and representing documents that societies see as worthy of preserving." In a more specific view, and paraphrasing Hjørland (2008) whose terminology is used in this thesis, KO is a "field of study" concerned with "*knowledge organization processes* (KOP)", which are "activities such as *document description*, *indexing* and

classification", done both by information professionals and "computer algorithms", through the use of "*knowledge organization systems (KOS)*" (p.86).

2.3.2 Knowledge Organization Systems (KOS)

The term KOS was coined by the Networked Knowledge Organization Systems Working Group (NKOS) in 1998. A KOS is considered mainly to be a tool for vocabulary control – this term is even used as a synonym of it (Leise, Fast & Steckel, 2003). KOSs systematize or arrange knowledge structures according to certain organizing principles. Such conceptual structures come from specific domains, mainly from the documents and terminologies that are produced and used by their "discourse communities", and its main use has been the description, indexing and classification of those documents for the purposes of retrieving the information contained in them, or the documents themselves. Hjørland (2006a) says that they are also referred to as 'semantic tools', because "[…] they are essentially selections of concepts supplied with information about their semantic relations to other concepts and symbols". The most common example of a KOS is a thesaurus.

As it can be seen in the concept map (p.36), this term is equivalent with Svenonius' term "bibliographic system", which doesn't have to be mixed with "information system". In Svenonius (2004) it equates to the term 'knowledge representations'. I will use this last term in this text to refer to those 'tools'. In this sense, it corresponds to the therm "library techniques" or 'LIS techniques' used for the Topic Maps community.

These three terms are problematic: for instance, the meaning of the term 'knowledge representation' has a different scope and coverage for the KR and Computer Science domain than in LIS, 'KOS' is not used by the Topic Maps community, and 'LIS techniques' is not part of the recognized terminology of the LIS discipline. I use them sometimes undiferentiatedly depending on the context, but I use mainly the acronym 'KOS'.

2.4 Conceptual models in LIS

2.4.1 Introduction

Conceptual models have been used mainly in database design, where they are called also abstract models (Carlyle, 2006). In Computer Science they implement the abstract

specification of data structures that define an entity-relationship representation of a domain. A conceptual model represents -but doesn't implement- 'concepts' (entities) and relationships between them.

The main conceptual models that were referred in the literature on Topic Maps in LIS are FRBR and CIDOC-CMR which will be briefly described next. One alternative proposal to the FRBR conceptual model is presented by Professor Shoichi Taniguchi, called "the expression-prioritized model". This is discussed in Berg (2004). Taniguchi (2003) also reviews the existing conceptual models in cataloging and other areas "such as archives, rights management, record keeping, and museums" (p.3).

Other conceptual models existing in the library domain are the Functional Requirements for Authority Data (FRAD), and Functional Requirements for Subject Authority Records (FRSAR). The Resources Description and Access (RDA) is the updated version of the Anglo-american Cataloging Rules (AACR2). It has been structured according to these and the FRBR model.

2.4.2 FRBR

FRBR (Functional Requirements for Bibliographic Records) is the best known conceptual model of library cataloging. It was developed by The International Federation of Library Associations (IFLA) from 1993 to 1998.

Since FRBR is a conceptual model, its main components are the *entities*, the *relationships* between them and their *attributes*. Entities, according to Carlyle (2006) "are things, either physical or abstract. Thus, an entity can be virtually anything: relationships are interactions among entities; and attributes are properties or characteristics of either entities or relationships." (p.266)

FRBR gives conceptual structure to those entities, attributes and relationships, and to the bibliographic records that register their bibliographic descriptions. Besides, it orders the terminology to refer to the entities in the "bibliographic universe". It was created with the "primary purpose of improving cataloging records (a product), cataloging (a process), and catalogs (a technology)." (Carlyle, 2003)

The entities conceptualized by FRBR are divided into three groups, as described by Tillet (2003) and defined in the FRBR final report (IFLA, 1998):

- Group 1: Work, expression, manifestation and item.
  - *Work:* "A distinct intellectual or artistic creation."
  - *Expression:* "The intellectual or artistic realization of a Work."
  - *Manifestation:* "The physical embodiment of an Expression of a Work."
  - *Item:* "A single exemplar of a Manifestation."
- *Group 2: Person and corporate body.* These entities are related to the entities in Group 1 by relations that show their roles with respect to the entities in Group 1.
- *Group 3: Subjects of works.* That is concepts, objects, events, places, and any of the Group 1 or Group 2 entities when they are subjects of works.
- *User tasks:* find, identify, select, and obtain. Additionally, FRBR recognizes the importance of being able to navigate.

Due to the scope of this thesis, FRBR will not be explained or studied in detail. However, which is interesting to observe for some of the conclussions, are some similarities with the concepts in Information Organization described earlier. With respect to 'document languages' and 'work languages', it would be easy to think of works and expressions (from group 1) and group 2 and 3, as entities to be described by work languages, because they describe information entities and not physical ones. Manifestation and item, in the other hand, refers to physical objects, and could be considered in the realm of what document languages describe. However, as it was explained in section 2.2.2.1, the distinction between information entities and physical entities is blurred. This has to do with the concept of 'bibliographic universe' and 'work' which is a difficult issue and the subject of ongoing theoretical discussions in LIS. For instance, a problematic issue in the context of this conceptual framework is the concept of "work" in FRBR, compared to the concept of "work" in Svenonius, but this issue is beyond the scope of the current thesis.

The implications though, for understanding the relation of FRBR with Topic Maps, are in the FRBR user tasks (which relate to the bibliographic objectives described by Svenonius – section 2.2.4), and in the fact that it is a conceptual model and thus a technology independent representation that can be flexibly implemented. The process of implementing it in actual catalogs has been called **FRBRization**. Aalberg (2005) explains that implementation can be

done at different levels: implementing the Group 1 (entities and inherent relationships), implementing the Group 2 and 3 (entities and inherent relationships), implement other relationships, and implementing the FRBR attributes.

2.4.3 CIDOC-CRM

The ICOM/CIDOC (The International Council of Museums) is an international organization devoted to the documentation of museum collections. The CIDOC Conceptual Reference Model (CRM) is an ISO standard (ISO 21127:2006) which provides a general data model for museums with the purpose to standardize enable information interchange and the integration of the museum community, as well as the integration of this community with other ones in the cultural heritage domain[12].

The CIDOC-CRM provides a reference ontology for the interchange of cultural heritage information, providing a description of the intellectual structure of data used in museum documentation as well as in other cultural heritage institutions: “This includes collections, sites and monuments relating to fields such as social history, ethnography, archaeology, fine and applied arts, natural history, history of sciences and technology.” (ICOM/CIDOC, 2009). Its purpose is to enable semantic interoperability by explaining the logical structure behind the structures currently used for description.

Since this purpose is similar to the purpose of FRBR in the library field, some efforts to integrate both conceptual models have been done. Aalberg (2005) claims that FRBR is “an ontology for exchanging bibliographic information within the domain and with other domains”. This could be, for example, the FRBRoo (object-oriented version, harmonized with CIDOC CRM), supported by DELOS, the Network of Excellence on Digital Libraries (Doerr & Le Boef, 2007).

The ICOM/CIDOC Reference Model is “the result of one of the most significant efforts for a formal representation of the basic notions of the cultural application domain.” (Amann, Fundulaki & Scholl, 2000, p.3). It is defined as “a high level ontology” by Stevenson and Styron (2006).

[12] http://cidoc.ics.forth.gr/

2.5 Topic Maps

This section is mostly based on Pepper (2008b) and on ISO/IEC 13250-1 and ISO/IEC 13250-2. Terms that are defined in the ISO standard are shown in bold.

As presented in the introduction, Topic Maps is an ISO standard (ISO/IEC 13250) whose idea began in 1991 to give a solution to the problem of merging the back-of-the book indexes of different computer system documentation by creating a superstructure over the documents.

Nowadays, Topic Maps is used in a variety of fields and purposes. Garshol (2007) summarizes them as Semantic Portals, eLearning, Business Process Modelling, Product Configuration, Information Integration, Metadata Management, Business Rules Management, IT Asset Management, Asset Management (Manufacturing).

In USA Topic Maps has been used by the Internal Revenue Service (IRS[13]) of the Department of the Treasury in USA (the "tax map"), the DOE (Department of Energy), Lexis-Nexis and in different E-Gov proceedings (Newcomb & Biezunski, 2003).

In Europe it is mainly used in the public sector, as well as in pharmaceuticals, automobiles, publishing and more (Newcomb & Biezunski, 2003). Its use is widespread in Germany and even more so in Norway. In this country there are numerous large and small scale projects using Topic Maps, some of which are listed by Garshol (2007): "forskning.no" (The Norwegian government portal to popular science and research information), "Kulturnett.no" (Norwegian public sector portal to cultural information), "Bergen Kommune" (City of Bergen citizen portal), "Apollon" (University of Oslo research magazine), "NRK/Skole" (for a curriculum-based browsing), among many others. Prof. Sam Oh (personal communication, April 12, 2009) expresses that one of the barriers in the dissemination of Topic Maps is precisely that the most successful cases are in languages that are not globally accessible. It should also be added that another barrier is in the fact that some of the applications are not intended to be visible.

[13] http://www.irs.gov/

2.5.1 Basic concepts

The basic conceptual building blocks of topic maps are topics, associations and occurrences. The model they belong to has been referred to as the TAO of Topic Maps (Pepper, 2000).

2.5.2 Topics and subjects

The Topic Maps concept of **subject** is anything (physical or abstract, real or fictional) that the topic map author wishes to make assertions about, i.e. assign a name, a property, or a role in some relationship with another subject. It is defined in ISO/IEC 13250-2 (the Topic Maps Data Model) as follows:

> "A subject can be anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever" (p.8)

A **topic** is the symbol or surrogate that represents the subject within a topic map. Topics are informally referred to as the 'proxies' of subjects in a computer domain (Pepper, 2008b). By definition, every topic represents a single subject. The goal of any Topic Maps application (often referred to as the "collocation objective") is to ensure that every subject is represented by one and only one topic.

Some topics are **typing topics**. A typing topic is "any topic that is used (or intended to be used) to type some other construct, whether it be a topic, association, association role, occurrence, or name." (Pepper, 2008b)

2.5.3 Identity

In order to achieve the collocation objective, Topic Maps encourages the use of explicit identifiers rather than names. Identifiers usually take the form of URIs (Uniform Resource Identifiers). These can be either **subject identifiers** or **subject locators**. and they are the basis for *merging*, a capability which is often described as the most powerful feature of Topic Maps.

Subject locators are URIs that identify subjects that are "network addressable information resources" and that have a location (an address) in an information system. The network addresses of such subjects can be used to identify them directly.

A subject identifier is a URI that identifies an arbitrary subject that may or may not have a location in an information system. It identifies its subject *indirectly* via a **subject indicator** (sometimes called a *subject descriptor*).

A subject indicator is simply a human-readable resource (i.e., document) to which a subject identifier resolves, and which is intended to convey the identity of the subject to a human. As the TMDM defines it, a subject indicator is an "information resource that is referred to from a topic map in an attempt to unambiguously identify the subject represented by a topic to a human being." Figure 1 exemplifies this 'indirect identification'.

Subject identifiers and indicators (or descriptors) can be "published", i.e. made available for use outside the scope of a specific application in order to achieve wider interoperability. They are then known as published subject identifiers (PSIs) and published subject indicators, respectively, or collectively as published subjects (Pepper, 2008b).

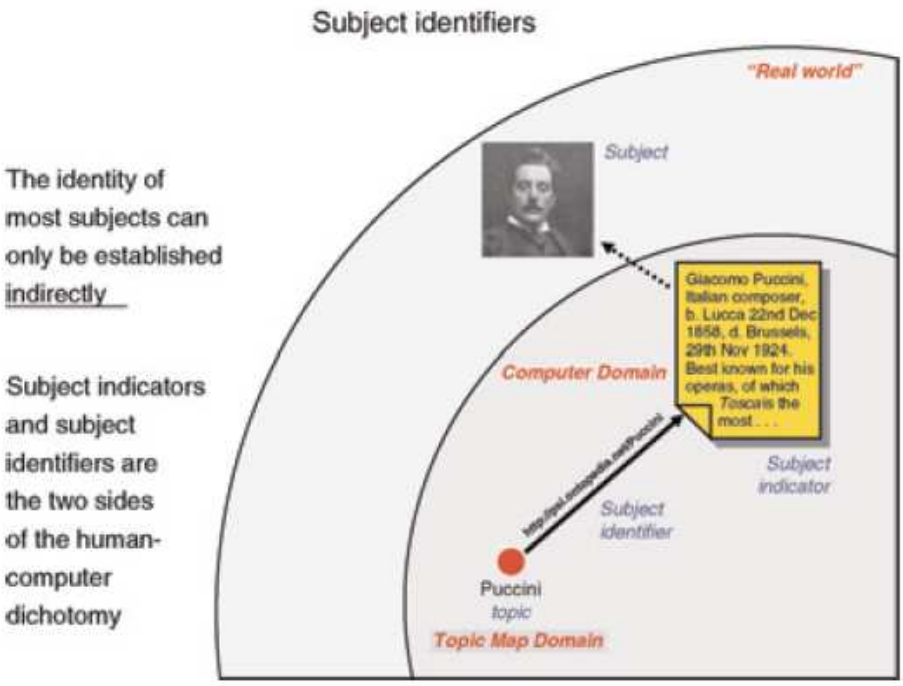

Figure 1: Subject identifiers and subject indicators (Pepper, 2008b)

The identification mechanisms of Topic Maps in relation to Information Organization will be described mainly in sections 4.1.1.1 and 5.2.1.

2.5.4 Names

Topic names are properties that have naming semantics. A topic can have multiple names, each of which consists of a **base name** and zero or more **variant names**. Each name is typed (i.e. assigned a name type) and may also be **scoped**. Figure 2. shows a example of these constructs:

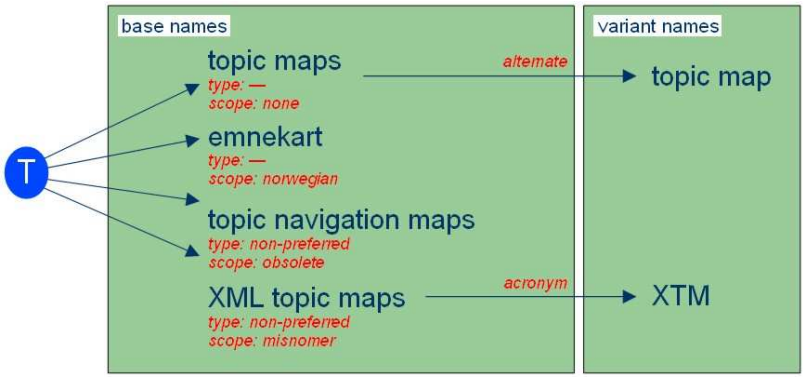

Figure 2:Topic names and variants (Pepper, 2008b)

A base name is the base form of a name, an alphanumeric string used as its default label.

Variant names are the alternative forms of base names that are optimized for particular computational purposes, such as sorting or display. The main cited examples of uses for variant names are sort key, plural forms, pronunciation, common misspellings/alternative spellings, and alternative orthographies.

2.5.5 Occurrences

Occurrences relate topics to relevant information resources that describe them. According to Pepper (2008b) the resource in question may be very small, such as a string representing a date, in which case it is normally included in the topic map and known as an “internal occurrence.” Or else it may be stored externally, because of its size, notation, provenance, or whatever, and referenced via a locator – normally a URL – which corresponds to a page number in a back-of-index (itself a locator for some piece of information relevant to the subject in question).

2.5.6 Associations and roles

Associations express relationships between subjects by relating one topic to (zero or more) other topics. They were originally meant to represent the ‘See also’ references that appeared in back-of-book indexes.

Each topic that participates in an association is said to play a role in the relationship that is expressed by the association. The nature of the subject’s involvement in a particular relation is expressed using a **role type**. e.g. Puccini plays the role of *pupil* in the teacher/pupil relationship with Ponchielli. This mechanism obviates the need for associations to have a specific direction, and all associations are therefore inherently multidirectional.

2.5.7 Scope

Scope is a set of topics that is used to qualify a statement (i.e., a name, occurrence, or association) with the purpose of indicating the context in which a certain assertion may be considered valid. If no scope is explicitly specified, the scope is said to be ‘unconstrained’. Topics that are used for scoping are informally referred to as ‘scoping topics’.

2.5.8 Merging

Merging is a process or operation and as such is different from the previous elements, which are constructs in the Topic Maps model. Merging can take place both within a single topic map (to eliminate redundancy) and when combining two or more topic maps. This process lies at the core of the Topic Maps paradigm, and can be traced back to the original motivation (merging indexes) that gave rise to the paradigm. While merging is an operation performed by an application, its procedures are strictly defined in the standard and it is based on the concept of identity described above.

2.5.9 Reification

Reification is the process of instantiating as a topic some Topic Map construct (a name, occurrence, association, role, or even the topic map itself) that itself is not a topic. Once this is done, whatever is represented by the construct in question becomes a subject in its own right, about which statements can be made. Reification is most often used to assign metadata to a topic map.

2.6 Knowledge Representation

2.6.1 General

According to Sowa (2000a), Knowledge Representation (KR) is an interdisciplinary field of study, derived as a branch of Artificial Intelligence (AI), which applies theories and techniques from logic, ontology and computation (Sowa, 2000a, p. xi-xii). The principles of KR, according to the same author are five: a knowledge representation is a surrogate, is a set of ontological commitments, is a fragmentary theory of intelligent reasoning, is a medium for efficient computation, and is a medium of human expression (Sowa, 2000a, p.134).

KR is closely related to simulation of human reasoning to model it in a way that computers can "understand", simulate it, and make inferences based on this. It is generally agreed that the main problem is that these processes take place inside human minds, and thus, their representation have to be based on things that only exist externally.

> "Any intelligent entity that wishes to reason about its world encounters an important, inescapable fact: reasoning is a process that goes on internally, while most things it wishes to reason about exist only externally. A program (or person) engaged in planning the assembly of a bicycle, for instance, may have to reason about entities like wheels, chains, sprockets, handle bars, etc., yet such things exist only in the external world." (Davis et al., 1993)

KR can take many forms and be applied in many fields and "things". In LIS for instance, it is applied to the objects of the "bibliographic universe", composed by documents which are at the same time representations of human creation and thought through the use of language (written, visual, graphic, acoustic). These documents are at the same time represented, through the use of a "bibliographic language" in "bibliographic descriptions" arranged in "bibliographic systems" or, "knowledge representations", as called by Svenonius (2004).

However, although some of the mentioned principles of KR could be applied to bibliographic languages and lead us think that those could be knowledge representations, the scope or purpose of KR compared to that of Information Organization and/or KO is broader and couldn't be considered as the same in those two disciplines. For instance, KR models events and operations (such as in the example presented by Sowa on the functioning of a system for

traffic lights), covers simulated behaviors (p.141), and tries to model logics and automated reasoning (p.4), using surrogates "to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it" (Davis, Schrobe & Szolovits, 1993), while bibliographic languages have a scope limited to the modeling of conceptual structures based on documents.

Perhaps what makes Information Organization and KO close to KR is the first principle of the latter one: "a knowledge representation […] is most fundamentally a surrogate, a substitute for the thing itself" (Davis, Schrobe & Szolovits, 1993), which could be a valid definition of *metadata.*

In Svenonius (2000) view, there are other kinds of knowledge representations nowadays, e.g. lexical databases, metathesauri, semantic networks, etc. (p.146). However, some of them have been developed outside the realm of LIS, and have been applied in the domain of KR and in other fields. Since Topic Maps transcends the LIS realm, some considerations found in the literature, relate it to these other knowledge representations looking at their possibilities in the LIS realm. Two of them are recurrent: semantic networks and ontologies. Here it is a conceptual explanation of both:

2.6.2 Semantic networks and ontologies

"Knowledge representation formalisms" are ways to represent knowledge graphically: mind maps, conceptual graphs, concept maps, semantic networks, are examples of them.

A semantic network or net is a graphic notation for representing knowledge in patterns of interconnected nodes and arcs" (Sowa, 1992). In this sense, Topic Maps is very related to them: "the basic model of semantic networks is very similar to that of the topics and associations" (Pepper, as cited in Colmenero, 2005). The main difference between semantic networks and Topic Maps, as explained in Pepper (2000, rev. 2002), is that Topic Maps adds the topic/occurrence axis to the topic/association model. Besides, like semantic networks, topic maps relate to knowledge representation formalis, but unlike semantic networks, they incorporate the ability to search.

The term **ontology** is used in KR, Computer Science and LIS with different meanings, which could be explained by the degree of formalization, their scope and purpose. In KR, ontologies are "the *categories* of things that exist or may exist in some domain […] a catalog of the types

of things that are assumed to exist in a domain of interest *D* from the perspective of a person who uses a language *L* for the purpose of talking about *D.*" (Sowa, 2000a, p.492). This definition is in accordance with what it's commonly agreed in the Topic Maps literature, where ontology is defined as the kinds of *topics*, *occurrences*, and *associations* that constitute a topic map (Pepper, 2000 rev. 2002).

The relations of this concept of ontology with AI, and the place of Topic Maps as a KR formalism won't be covered in this thesis. But some clarification is needed: in KR there is a difference between 'terminological ontologies' and 'axiomatized ontologies', as well as between 'informal ontologies' and 'formal ontologies'. Their distinction is based on their degree to support complex inferences and computations according to their axioms and definitions, and in their use of language, from natural language to formal languages that can be translated to logic (Sowa, 2000a, p.493).

Krishnamurthy, Müller, Kenny, and Sternberg (as cited in Yi, 2008) consider ontologies as a "way of making the concepts within a domain, as well as concepts across diverse domains, and the relationships between these concepts machine processable." (p.1900). This meaning is commonly assumed in Computer Science.

In the LIS community there seems to be a common understanding of ontologies as thesauri where the relationships (mainly hierarchical) are extended (or named). This topic is extensively examined in the KO literature. Fischer (1998) associates them to the concept of "terminological ontologies".

There is also a classification of ontologies that the Topic Maps community has paid attention to: the domain related ontologies (with a tag focus and a specific scope) and encompassing ontologies, or metaontologies. (Newcomb & Biezunski, 2003)

Svenonius (2000) considers ontologies in the same sense that is assumed in Topic Maps: "a theory regarding entities, especially abstract entities to be admitted into a language of description" (p.31). Since all topic maps projects are related to ontologies, this common perspective between the two communities is the one that is going to be assumed in this thesis.

2.7 A common conceptual framework

The following concept map presents the main concepts that have been described in this section. Concept maps are not related to Topic Maps. A concept map is a graphical tool for organizing, representing and presenting knowledge (a knowledge representation formalism).

In this concept map, the concepts tied with a defined line come from the framework described above and were linked according to my interpretation of the authors that described them. Svenonius' main concepts are marked with a stronger line. The concepts taken from KO discipline are in a grey background. The concepts in circles with pointed lines are mainly terms that emerged from the literature (Appendix 1) and that I located in the main framework according to my interpretation.

Each section marked with a circle and a letter corresponds to a chapter, either in the conceptual framework, the literature review or the synthesis chapters, or in the three of them. There is a table at the bottom of the concept map indicating them.

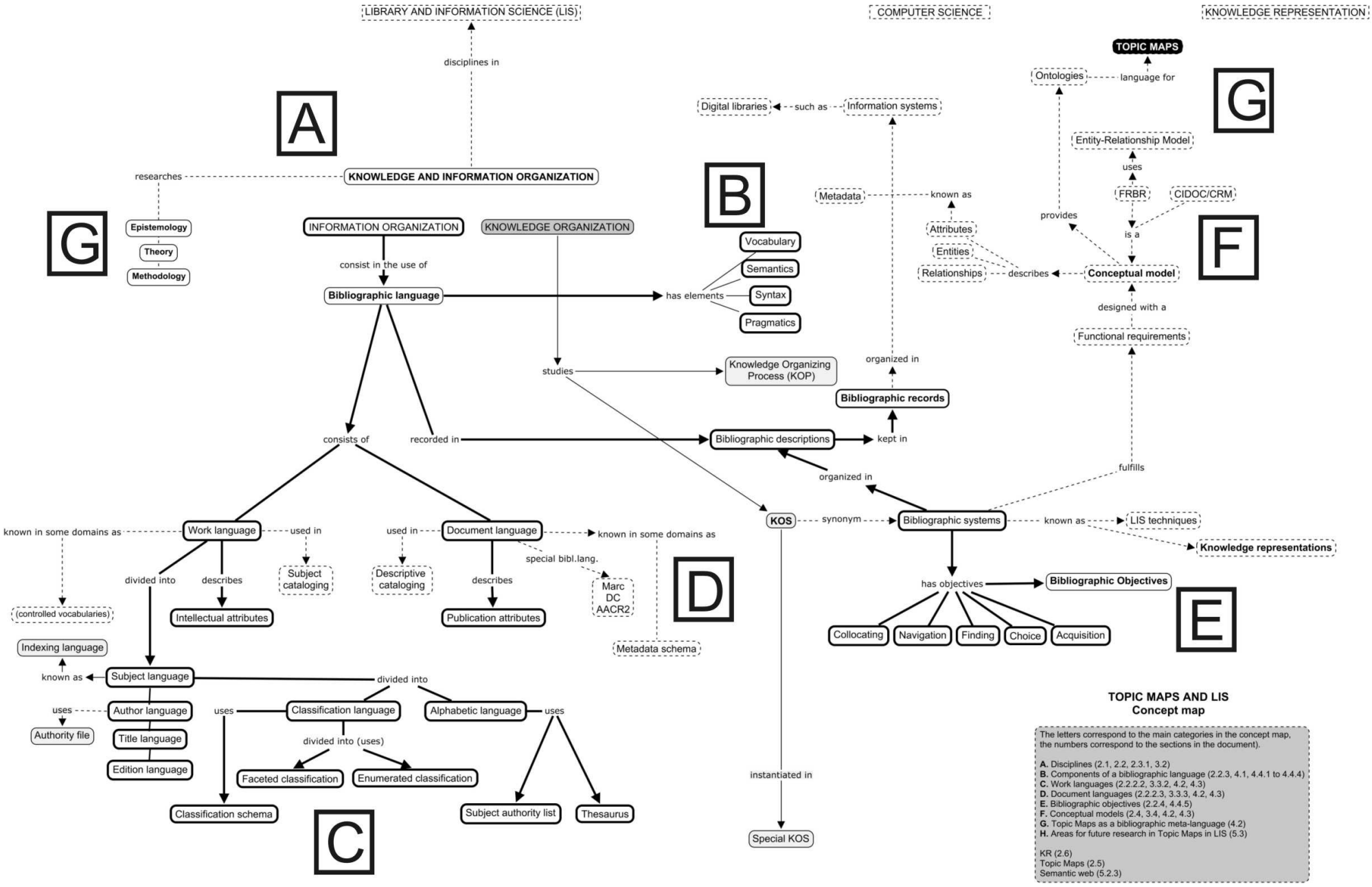

Figure 3: Concept map of Topic Maps, Knowledge Organization and Information Organization Concept

3. Literature review

This is a literature review on Topic Maps in LIS. The organizing principle and the terminology used on it corresponds to the concept map on page 44[14]. In the corresponding reference list, each of the sources cited in this chapter that are considered relevant for a bibliography on Topic Maps in LIS appears with an astherix.

3.1 Introduction: Topic Maps in the LIS community

One of the factors indicating the adoption of a concept into a discipline is its appearance in publications, conferences, international bodies' plans and institutional projects. In the LIS literature and community, Topic Maps started to be a focus of interest around the year 2000, when it became an ISO standard. The Topic Maps community, on the other hand, acknowledges its roots in the Information Science tradition and has been interested in finding the connections of Topic Maps to the LIS concepts and practices. These concerns and the practical experiences in applying Topic Maps in that field have been presented in conferences, blogs, and mailing lists.

Topic Maps has been discussed in journal articles of long tradition and importance in the LIS community, among others the Journal of Information Science (Garshol, 2008), the Journal of Library and Information Science (Peng & Ke, 2008), the Journal of The American Society for Information Science and Technology (JASIST) (Yi, 2008), Information Research (Tramullas & Garrido, 2006), OCLC Systems & Services (Stringer-Hye, 2005) and Library Hi Tech (Iglesias & Stringer-Hye, 2008). It appeared for the first time in the Annual Review of Library and Information Science in its 2007 edition, as part of a section on 'ontologies on the Semantic Web' (Cronin, 2007, p.430).

Besides, Topic Maps has been presented to the LIS community at conferences such as the International Symposium for Information Science (ISI), The European Library Automation Group (ELAG) conference, the International Association of Technological University Libraries (IATUL) conference, the Digital Library Federation Spring Forum, ECDL, the International Association of Music Libraries, Archives and Documentation Centres (IAML)

[14] From here, all the Topic Maps-related concepts are italicised and the LIS and related terminology is used according to the conceptual framework (Chapter 2).

conference, and the Digital Resources for the Humanities (DRH) conference. The connections of Topic Maps with LIS have been introduced to the Topic Maps community at the International Topic Maps Users Conferences in Oslo, various XML related conferences, Topic Maps Research and Applications (TMRA) in Leipzig, and The Australian World Wide Web Conference.

Topic Maps was also presented at the series of conferences 'Luminary Lectures at Your Library', organized by The Library of Congress (Newcomb and Biezunski, 2003), where the authors explained to librarians the main principles of Topic Maps as well as their experience in applying it to the "tax map": master indexes to tax publications. Topic Maps was also introduced at LITA 2006 National Forum, where Steven Newcomb and Patrick Durusau, co-editors of the Topic Maps reference model, presented what Topic Maps is in relation to the vision of subject-centric computing (Newcomb, 2006). An interest group for Topic Maps originated from this forum, but it doesn't seem to be currently active. A "Topic Maps awareness seminar" took place at the National Library of Australia in 2004 (Johannesen & Pearce, 2004).

Some institutions and organizations in the LIS community have considered Topic Maps for possible implementations and have conducted evaluations of its feasibility: The Aquifer initiative of the Digital Library Federation (DLF), whose purpose is to promote the effective use of distributed digital library content for teaching, learning, and research in the area of American culture and life[15], conducted a survey to gather information about uses of digital collections in DLF libraries. In their 2006 report, Topic Maps was in a list of possible technologies that could solve the challenges in navigating digital collections and locating objects among them. OCLC (Online Computer Library Center) claims to use Topic Maps in two of its projects: "RDF Topicmaps", which "explores subject navigation of Web sites using semi-automatically generated finding aids"[16], and "The WordSmith Project", which developed a software for extracting concepts automatically and organizing them into subject hierarchies (Godby, C. J., Miller, E. & Reighart, R., 2001); however, these projects are based on RDF and not on Topic Maps. In their annual conference in Seoul, IFLA (the International Federation of Library Associations) included a paper on "The Living Memory" project (Leuenberger et al. 2006), an application of Topic Maps to the image domain. Finally, at the

[15] http://www.diglib.org/

[16] http://www.oclc.org/research/projects/rdf_topicmaps/

National Archives and Records Administration (NARA) in the United States, Topic Maps was studied as one possibility for searching their archive collections (Le & Nguyen, 2007).

The use of Topic Maps for e-government, education, information architecture of corporate sites and publishing in Norway is widespread both in private companies and in the public sector. In 2002, the first conference on Topic Maps in the world took place in Oslo, and it is also the country where most editorial companies have adopted Topic Maps (Vogt, as cited by Kongsbakk, 2004). The library sector in this country has shown an interest as well. The ABM-Utvikling (the Norwegian Archive, Library and Museum Authority) funded a project where Topic Maps was used for the FRBRization of BIBSYS, one of the main national library catalogs, as part of the development of the Norwegian Digital Library Initiative (Kongsbakk, 2004, p.4; Aalberg, 2005). This project cooperated with OCLC.

Some national libraries such as The Royal Library in Denmark (Laursen, 2006), the National Library of Korea (Oh, 2007) and the National Library of Hungary (Tóth Máté, personal communication, April 19, 2009), The National Library of New Zealand (in what relates to the NZETC project) have also considered or applied Topic Maps in various ways. Some of the details will be explained later.

Finally, the placement of Topic Maps in LIS has been the exclusive topic of one mailing list: TopicMaps in LIS[17] and a wiki project (Topic Maps for Libraries, 2006).

Concerning the relation between the discourses and communities, the Topic Maps and the LIS community, some authors have shown the need for interdisciplinary discussions on Topic Maps, which can act as a "catalyst" (Sigel, 2000, p.1) or as "a means of 'bridging the gap'" (Pepper, as cited by Colmenero, 2005) for the integration of perspectives of different but related communities: LIS and Computer Science.

Adams (2002) insists that the library expertise in Information Organization (building taxonomies for instance), as well as that of computer scientists (developing ontologies), is needed to achieve the idea of the semantic Web (see 5.2.3); and further claims that Topic Maps is one of the technologies that could make this vision possible. In the same line, Stringer-Hye (2005), after the conference XML 2000, where the idea of the Semantic Web was presented by Tim Berners-Lee, advocates the need for participation from the library

[17] http://ligent.net/mailman/listinfo/topicmapsinlis

community with their expertise in the construction of such a vision. Pharo (2004) also expresses his opinion about how the library community (specifically the Knowledge Organization area) should take Topic Maps as "an opportunity to implement well-known knowledge organisational principles and to try out principles that are less hierarchical" (p.11). Also Sigel (2003), by showing that KO and Topic Maps are complementary and overlap in many aspects, concludes that Topic Maps has an impact in both communities and calls for the observation of practices and research among them (p.384).

3.2 Topic Maps and Knowledge Organization

The studies of Topic Maps from the point of view of the LIS concepts and theories have been done primarily from the perspective and concepts of the KO discipline. Sigel (2000) was one of the first authors to present this vision. In a conference paper at XML Europe, he gave a critical look at the possibilities that Topic Maps offer in addressing one of the main problems in KO: "how to reorganize, enhance and semantically integrate heterogeneous subject data" (p.1). Sigel also sees that one of the main potentials that Topic Maps has for KO is its capability to allow decentralized creation and exchange of metadata. This entails a challenge to "redesign" KO methodologies for "collaborative knowledge building activities on distributed resources" (p.1). Sigel sketches "three typical scenarios in which heterogeneous metadata occur", showing how the same problems in KO appear even using Topic Maps, but explains why Topic Maps presents valuable options to solve them. These problems are: mapping, merging and integrating. His main conclusion is that

> "You should seriously consider Topic Maps if you plan for applications on top of digitally organized intellectual assets. But make sure to include strong KO expertise in your [knowledge management] methodology and team applying Topic Maps, since the classical challenges of KO will inevitably haunt you in a new disguise, even with this promising technology." (Sigel, 2000, p.2)

Since 2000, these topics have been recurrent in the literature: Pepper et al. (2000) showed the possibilities and limits of the Topic Maps paradigm and its application for practical 'information work', looking at the potential strengths and weaknesses of Topic Maps to cope with the challenges in KO, and how Topic Maps relate to other proposals in the area of semantic knowledge markup/semantic Web. Colmenero (2005) described the possibilities that the Topic Maps "metamodel" presents for knowledge organization, semantic navigation and

interoperability. Sigel (2003) wrote a chapter for "XML Topic Maps", the most comprehensive book on Topic Maps. There, he reasumes some of the topics described in 2000, such as the relation between KO and Topic Maps in both directions: relevant theory and practice in KO that should "inform TM design" (p.425), and the impact of Topic Maps in the theory and practice in KO, looking at "how Topic Maps might aid the process of organizing knowledge, what this might mean in practice, and which consequences this might have for indexing theory" (p.384). He found that Topic Maps can aid KO in achieving one of its main goals which is semantic interoperability, by allowing "flexible indexing views", "scope filtering", "semantic retrieval", and "ontology-based modeling". In this way it would be possible "convert structure medatada and indexing language into knowledge networks" (p.?). On the other hand, KO can contribute to Topic Maps by bringing a long expertise and solid principles and methods for the solution of problems of organizing knowledge that may arise in Topic Maps as well. For example, one of the capabilities of Topic Maps is *merging*, which is the basis for semantic interoperability, but KO expertise is needed to achieve it successfully.

The conclusion about the relation between Topic Maps and KO seems to be that "Topic Maps is instrumental to KO, since they constitute (yet another) basic technology which KO may employ inter alia within a broader methodology in order to provide improved information retrieval and collaboration services." (Sigel, 2000, p.3); that is, Topic Maps is "a new enabling technology for KO" (Sigel, 2003, p.425) in the sense that it provides both a model and a syntax (XML) to represent and exchange the products of KO. KO is more interested in the "intellectual and social processes" that make the modeling and semantic integration of conceptual structures of different discourse communities possible.

These "discourse communities" sometimes become silos with disparate terminologies and solutions for the same concepts and problems. Since Topic Maps originated in a different community than LIS, some of the terminology doesn't seem to correspond with that of the LIS tradition. Sigel (2006) tried to map these terminologies, as well as Hjørland (2006b), who presents a critique of Topic Maps, saying that its terminology "seems quite idiosyncratic". Mapping terminologies is an issue related to the interdisciplinarity of Topic Maps, which is, according to Adams (2002), a need to encourage collaboration, a concern that is indeed being reflected and studied by the Topic Maps and LIS communities.

In this respect, both Sigel (2006) and Hjørland (200b), as mentioned, agree in equating some of the main concepts of Topic Maps to already existing concepts in KO: *topics* to 'concepts',

associations to ‘relations’ and *occurrences* to ‘information resources’ or ‘documents’ in the KO terminology. Additionally, Hjørland (2008b) states that *topic types* would equate to ‘categories’, and *occurrence roles* to ‘document types’. This, however, is something that requires a more detailed exploration (section 4.1).

For instance, the concept of ‘subject’, which is defined clearly defined in Topic Maps (section 2.5.2), has been problematized in LIS by Hjørland (1992). Although the concept of *subject* in Topic Maps deserves an epistemological exploration, for the purpose of this thesis, I will use a basic differentiation, proposed by Fugmann (as cited in Sigel, 2006), in which he asserts that a ‘subject’ is “everything one can make a meaningful statement about”, and a ‘concept’ is the “sum of all essential statements which can be made about a subject”. A name would be, in Fugmann terms, the a linear string designating (naming) a subject or concept.

Besides, Hjørland (2006b) and Pepper et al. (2000) note that the term *facet* as originally used in Topic Maps conflicts with its use in ‘faceted classification’ in the KO realm. However neither the term nor the concept is part of the latest version of the ISO standard.

Likewise, Hjørland (2006b) finds that the term *theme* has a completely different meaning in the Topic Maps terminology than in LIS. In Topic Maps, *theme* was used to denote a member of the set of topics used to specify a *scope*. However, the term was jettisoned in recent versions of the standard in favor of *scoping topic*.

3.3 Topic Maps and bibliographic languages

3.3.1 General

One of the main concerns found in the literature on the placement of Topic Maps in LIS is their relation to bibliographic languages: if they are one of them, if they evolve from them, or in which way they relate to the existing ones.

Colmenero (2005) describes their similarities and the possible applications of Topic Maps to them in digital libraries. Garshol (2004) studied and compared the Topic Maps model with the other bibliographic languages (including, as complement to the previous authors, document languages such as DC). Garshol states that since Topic Maps was born from the need of merging electronic indexes, it belongs to the tradition of subject languages (p.385). Chen & Chen (2001) went a step further, studying Topic Maps in comparison with different systems

of classification and cataloguing, but also including search engines, data mining, and semantic networks. Pepper (2008b) presents an overview of several of these previous approaches.

In this context, Topic Maps has been defined as as a “metadata format” (Walsh and Dalmau, 2006), but mostly as a “metamodel” (Marzal et al., as cited in Colmenero, 2005), a “meta data model“(Johannesen, 2007), a “generalized data model” (Johannesen, 2006), “both a conceptual model and an XML exchange format” (Johannesen & Pearce, 2004), an “ontology framework for information retrieval” (Garshol, 2004) or simply “a framework” (Johannesen, 2007). Topic Maps was presented as well in a conference called “Metastructures” in 1999. More informally, Topic Maps has been designed by one of its creators as “neutral envelopes, hospitable to any existing or future schema for knowledge representation” (Biezunski, 2003, as cited by Sigel, 2003, p.435). Pepper (2008b) explains this as the main “value proposition” of Topic Maps:

> “One of the most important aspects of Topic Maps is the ability of the model to represent virtually any kind of knowledge structure or data model. It is this capability – coupled with the ability to merge arbitrary topic maps – that underlies its value proposition: improved information management and enhanced findability through connecting disparate systems and collating information and knowledge from different sources”. (p.23)

It seems then that the main relation of Topic Maps to bibliographic languages is its representational capability. Johannesen (2006) says that, because of this characteristic, “you can do any classification scheme and structure inside Topic Maps”. And Ahmed (2003) proposed the use of “design patterns” (a term used in Computer Science) to represent bibliographic languages, based on the generalizations that can be made based on solutions that are recurrent in topic maps design.

The way to apply such representations to KOS (i.e. taxonomies, thesauri, faceted classifications, metadata schemas and hierarchies) is what is reviewed next. The terminology of the authors was adapted to the conceptual framework and the synthesis is proposed in chapter 4. For clarifications on these concepts refer to chapters 4 and 2.

3.3.2 Work languages

3.3.2.1 General

As it was stated, Topic Maps has been considered by some authors to be one of the "evolutions" or developments of KOS: Sigel (2006) describes Topic Maps as one stage in the development of the semantic technologies (glossaries, folksonomies, taxonomies, thesauri, topic maps, ontologies). This author illustrates the evolution in the following diagram:

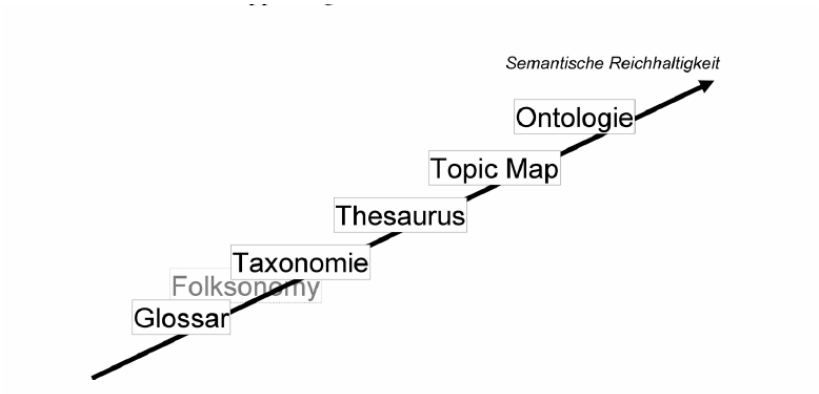

Figure 4: The "Semantic Ladder"
(Blumauer, A., & Pellegrini, T., 2006, cited in Sigel, 2006).

Garshol (2004) sustains and explains this same idea, arguing that Topic Maps "go beyond the traditional solutions" by making possible –not only to represent them all, but to extend them in ways that improve information retrieval (p.378).

This capability to model and extend existing work languages has been recognized by Colmenero (2005); Garshol (2004); Johannesen (2006); Pepper et al. (2000); Pepper (2006); Pepper (2008b); Pharo (2004); Sigel (2003); Oh (2007); Oh (2008b); Oh (2008c). This fact seems to be the key issue of the role of Topic Maps in LIS.

3.3.2.2 Subject languages

Classification languages

As it will be explained later (chapter 4.1.3.3.) classification schemas have a taxonomic underlying structure which uses a hierarchical relational semantics. The kinds of representations that the authors have proposed usually point in the direction of the representation of these particular structures and semantics, not in the representation of

particular cases of KOS, such as the DDC, UDC or Bliss. This, as is pointed out later (chapters 4.2 and 5.3) is a problematic area rich in possibilities for future research.

Enumerated classification languages

Taxonomies of the type 'genus-species' have been modeled in Topic Maps. Ahmed (2003) described with two design patterns the hierarchical naming and the hierarchical classification, especially of the mentioned type 'genus-species'. A topic represents a class in a bibliographic language. The hierarchy is represented by an *association type* (named "Hierarchical Relation Type") which relates two classes (*topics*), playing either the *roles* of subordinate or superordinate. The elements that belong to a class are represented as well as *topics* which participate in an *association type* with the *roles* of "classified as" or "classification". The items or subjects that are classified under a classification scheme will be associated with its class using the role type "instance". Ahmed (2003) also presented a "hierarchical naming design pattern" which provides a solution to the problems that may arise with the "Topic Naming Constraint"[18], that is, when *topics* representing different categories were merged based on their names. It also makes possible the display of a complete name of a *topic* which shows its place in a hierarchy (a kind of breadcrumb), and a short name that would be used when the context or place in the hierarchy is already shown.

Faceted classification

Faceted structures are together with thesauri one of the relational semantic structures most often explored in the literature. This is due, perhaps, to the fact that some of the main applications of Topic Maps are in the domain of Information Architecture, where faceted structures are commonly used.

Faceted structures are applied in many different domains. In a KOS perspective, a faceted structure is the underlying structure of a 'faceted classification'. The later one is considered as a "scheme of bibliographic classification based on the analysis of subjects according to a set of fundamental concepts, usually personality, matter, energy, space" or as classification schemes "whose terms are grouped by conceptual categories and ordered so as to display their generic relations" (Feather & Sturges, 2003).

[18] This "Topic naming constrain" has disappear in the current version of the Standard. It prohibited a processed topic map from containing multiple *topics* with the same base name in the same scope.

To represent faceted structures in Topic Maps, Pepper (2008b) proposes a solution using "a set of *topics*, each of which is related via an association to a *topic* that represents a facet, and some of which may be organized hierarchically via additional *associations*".

Garshol (2004) uses an existing ontology (XFML, see below) which extends the ontology of a thesaurus "by adding a new type 'facet' and a new *association type* 'belongs-to- facet'" (p.387). In this model each facet will be represented by a *topic* of the type 'facet' with the name of the facet as a *base name*. Two association types connect both the top level term of each facet (its class) to its facet (belongs-to-facet' *association type*), and the terms below the top level term (subfacets or arrays) to it through a broader/narrower (BT) *association type*.

Ahmed (2003) presents a design pattern for faceted structures. In his model, each facet is treated as a separated hierarchy (for which he also provided a design pattern –see above-). The pattern represents each facet with a *topic*, using its name for the facet. Besides, there is also a *topic* to represent the class. The types of *associations* both between the facet and the class, and the subfacets (or arrays) with the facet, define the type of hierarchy to use and are the main *association types* to use. This is the main difference with Garshol (2004), who only considered the BT *association type*. The following is an example of Ahmed's design pattern applied to a classification schema:

```
    <baseNameString>Duboeuf 2001 Cabernet Sauvignon</baseNameString>
  </baseName>
</topic>

<!-- Wine's classification on the Wine Type facet -->
<association>
  <instanceOf>
    <topicRef xlink:href="#wine-of-varietal"/>
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#varietal"/></roleSpec>
    <topicRef xlink:href="#cab-sauv"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#wine"/></roleSpec>
    <topicRef xlink:href="#duboeuf_cab_2001"/>
  </member>
</association>

<!-- Wine's classification on the Region facet -->
<association>
  <instanceOf>
    <topicRef xlink:href="#wine-from-region"/>
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#wine"/></roleSpec>
    <topicRef xlink:href="#duboeuf_cab_2001"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#region"/></roleSpec>
    <topicRef xlink:href="#burgundy"/>
  </member>
</association>
```

Figure 5. Sample of XTM file of a faceted classification (Fragment of an XTM file of a faceted structure in Topic Maps (Ahmed, 2003)

Garshol (2004) and Colmenero (2005) reference the eXchangeable Faceted Metadata Language (XFML), which is a XML format for sharing faceted classifications developed by Peter van Dijk (2003). It provides both an ontology for faceted classification and the interchange format in XML, which can be converted, using an XSLT style sheet into XTM (a topic map) (Garshol, 2004).

Alphabetic languages

Regarding the studies or views relating Topic Maps to specific KOS, Kongsbakk (2004) conducted a detailed study on the similarities and differences of Topic Maps and **thesauri,** both from a theoretical perspective and from a practical one. She found the main differences between the two in eight aspects. I put in parenthesis some sections in this work where her findings are also described:

- *Their background:* The thesaurus has existed before the Web and has its roots in the Knowledge Organization tradition, while Topic Maps has developed in the environment of the system developers. (Introduction and chapter 5.2.3)
- *Their function and purpose:* The purpose of thesauri is to serve in the indexing and searching processes, as part of an information retrieval system; while Topic Maps is an information retrieval system in itself because it includes searching and navigation capabilities, and relates instances to the topics and not vice versa. (See chapter 4.4.5.1).
- *Their structures:* Thesauri are mostly hierarchical while Topic Maps is based mainly in a networked structure. (See chapter 4.1.3.3).
- *Their subjects:* What is considered a subject in thesauri and Topic Maps is different. For the first one they are concepts related to what the documents are about, for Topic Maps everything (document related or not) can be a subject. (See chapter 3.5).
- *The relationships:* In Topic Maps there are no limitations when it comes to relationships while in thesauri there are only three ways to represent them (equivalence, hierarchical and associative (2003, p.26). (See 4.4.2).
- *The linguistic treatment:* Thesauri treat linguistically their terms through equivalent relations and qualifiers. Topic Maps allows expressing perspective and identity. (Chapters 4.1.1. and 4.4.4).

- *The coverage of the standards:* Thesauri standards focus on vocabulary control and vocabulary building and are mostly based on nouns. Topic Maps standard doesn't specify any vocabulary, allowing either nouns or verbs to be used, and covers technological implementation to ensure exchange of information between computers.

Beltran et al. (2002) discuss the use of verbs in thesauri presenting Topic Maps as the enabler model for this treatment.

One of the main conclusions regarding the representation capability of thesauri by Topic Maps is that it allows 'extending' the fixed relations of thesauri. Those ones are limited by the vocabulary that names them to: BT, NT, RT, UF/USE, and sometimes TT.

Among others, Yi (2008) and Garshol (2004, p. 386) agree on the fact that almost all the possible relationships between terms can be described and/or made explicit with Topic Maps:

> "Perhaps the most powerful aspect of Topic Maps is the ability to create associations between topics. Using Topic Maps, almost any association can be described […]. Associations can explain relationships like metadata "describes" data." While thesauri cannot show explicit relationships to users, Topic Maps-based ontology defines relationships and shows these relationships explicitly" (Yi, 2008).

Some authors have shown how to represent thesauri with topic maps: Garshol (2004) and Ahmed (2003) represent the two views or possibilities to do so: a term-centric approach (explained by Ahmed), and a concept-centric approach (explained both by Ahmed and Garshol).

1. Modeling terms as names (a term-centric approach).

Ahmed (2003) calls this model "Topic Per Concept Thesaurus Pattern" and describes it as follows: each of the terms of a thesaurus (independently if they are authorized terms or not) are represented as topics in a topic map, using the term string as a base name of the topic, and each term at the same time being the "subject" of the topic. The hierarchical and equivalence relations are expressed as binary associations between the correspondent terms. The notes are represented as occurrences and their types (scope note, history note, translation note and so on) are made explicit through the reification of each specific occurrence. The related term associations (RT) are represented in this approach as "role types" of the "term relationship"

association, which is limited by Ahmed (2003) in his description and PSIs to the "part-whole" relationship (he also includes as a "term relationship" association the other relationships among terms including the hierarchical one, and the equivalence one as a result of the term based representation, that is taking each term as a topic, the USE/UF relationship is presented as an association, which doesn't happen in the following model).

2. Modeling terms as subjects (a concept-centric approach)

Ahmed (2003) calls this model "Topic Per Term Thesaurus Pattern". Pepper (2008b) and Garshol (2004) also presents this approach. Garshol (2004) explains it as follows: each term that has no USE relationship is represented as a topic of type 'term', and used as the base name of the topic as well. The terms that have a USE relationship are represented as names of the preferred topic but using scope to specify that they are a 'non-preferred term'). Scope notes are, as in the previous model, represented as occurrences, but no reification is used to specify this. Instead, an occurrence type is used. The related term (RT) relationship is represented as an association of type 'related term'. The BT and NT relationships are represented as an association of type 'broader/narrower' (with the roles specifying which topic is broader and which is narrower). Ahmed (2003) uses the same approach in his specification of the design pattern, creating one topic for each single concept, allowing it to have multiple base names according to the equivalence relationships, making use, as Garshol (2004) did, of the scope functionality to show when a term is "non-preferred". This model eliminates the associations used to relate synonyms in the previous approach because it makes use of the multiple names instead. Ahmed observes that this model, compared with the previous one, has the limitation of not allowing to add notes to individual terms but only to the concepts (the preferred term), but has the advantage of creating a "much more compact topic map and also one which is easier to process for display purposes, as one need only enumerate all of the names of a topic to list all synonyms rather than follow associations".

These two approaches are, in general terms, valid for representations of any kind of KOS.

Ontologies

Although including ontologies as "alphabetic languages" sounds problematic (see chapter 2.6.2), the capability of Topic Maps in extending relations makes ontologies a common topic of discussion in the framework of Topic Maps and alphabetic languages, such as thesauri. According to Garshol (2004, p.386) and Yi (2008) Topic Maps shares with ontologies the

characteristic of using “open vocabularies”, which provides them with flexibility in representing any kind of relationship between the terms.” Ontologies (in the LIS perspective of the term explained in 2.6.2) have also been modeled with Topic Maps. For instance Sigel (2006) reports on modelling the Integrative Cross-Language Ontology (ICLO) with Topic Maps (Schmitz-Esser & Sigel, as cited in Sigel, 2006)

3.3.2.3 Author languages

Johannesen (2008) jubilantly observed about Topic Maps and authority records: “shared global authority control (finally a model for this!)”.

However, in the literature there were no descriptions of representations of author languages in Topic Maps. Norrish & Stevenson (2008) report on the creation of a system called EATS (Entity Authority Tool Set) in the framework of the NZETC which uses Topic Maps.

It is important to notice though that author languages are, from the point of view of LIS a very important type of bibliographic languages, due to the kinds of entities they control (the authors!). But from the point of view of Topic Maps they are term lists that support equivalent relationships (see chapter 4.1.3.3).

(Almost) every individual topic map project has included the *topic type* ‘person’, and thus, has in a *topic map* way modeled author languages. The main concerns, though, of this kind of modeling are related to the provision of “unique, persistent, sharable identifiers for any sort of entity” (Norrish & Stevenson, 2008). Identification of entities is considered as one of the prerequisites for authority control (see section 4.1.1.1).

The EATS model, for instance, was built taking into account the existing systems for authority control in the library world, while also aware of the limitations of the model underlying those systems, which were considered “inadequate for a digital world”. The idea of EATS dates back to 2006, when the NZETC used MADS for the automated mark-up of references to topics of interest. MADS, as Norrish and Stevenson explain, is a Library of Congress standard defining an XML schema for an authority element set to provide metadata about people events, and topics.

3.3.2.4 Title languages

The same observations made on the representation of author languages by Topic Maps apply to title languages, with the additional fact that *names* and *variant names* would be the concern of this kind of modeling. Some authors report on the use of these Topic Maps constructs, but they were treated in other sections (chapter 4.1.1.2; 4.4.1 and 4.4.4).

3.3.3 Document languages

Document languages seem to receive less attention than work languages in the literature on Topic Maps in LIS, especially from the LIS community. However, some authors have shown that metadata schemas are not only possible to be represented, but to be enhanced and expanded with Topic Maps.

3.3.3.1 Dublin Core

Pepper (2008a) and ISO/IEC DTR 29111: 2007 present proposals for expressing Dublin Core using Topic Maps.

Pepper (2008a) summarizes the previous proposals to do so, which are represented by Algermissen, Pepper, and Maicher (as cited in Pepper 2008a).

- Algermissen uses *names, associations* (typed -as in 'creator'- and untyped ones -as in 'relation'-), and *occurrences* (at different levels: for the resource and for the properties that describe the resources) to handle 14 of the 15 core elements.
- Pepper represents the property 'title' as an untyped *topic name,* 'description', 'date' and 'rights' as *occurrences,* 'DC identifiers' as *subject identifiers* for *occurrence types* and *association types*; and *associations* for all the other elements.
- Maicher has a similar approach to the previous one, but covers all four sets of terms documented in [DCMI Terms].

The proposal of Oh et al. (2007) -which is an ISO draft-, specifies how to represent in Topic Maps each of the elements of the DC Metadata Elements Set:

- *As association types:* contributor, coverage, creator, format, language, publisher, relation, source, subject, type
- *As occurrence types:* date, description, identifier, rights
- As names: title

Besides, Oh et al. (2007) also include a suggestion on how to represent DC Other Elements and Element Refinements, and a suggestion for representing both the Encoding Schemes and The DCMI Type Vocabulary. This could be done by representing each of the~~ir~~ terms as *topics* "whose subject identifier is the same as the URI of the term in question".

Pepper (2008a) generalized this conversion making an abstract procedure, which is based on the idea that "assigning metadata to resources is equivalent to making statements about *topics*. It is therefore natural to represent the assignment of property/value pairs as statements of various kinds". In the case of the two DC vocabularies that define metadata elements (and element refinements) the key issue, in his words, is to decide whether to represent a given property as a *name*, *association*, or *occurrence*. (See chapter 4.2)

Garshol (2004) defines DC as a vocabulary for metadata: "The Dublin Core specification defines the meaning of each property, but is silent on how to represent both the properties and their values, and is thus independent of any particular technology."

Tramullas & Garrido (2006) actually seem to have applied the representation of DC into Topic Maps in the development of Potnia (chapter 3.6.4). In the research done for developing that application they report to have used Topic Maps as the framework for the Dublin Core metadata set using XTM.

3.3.3.2 MARC

"The Machine-Readable Cataloging (MARC) standards have been widely used for the representation and interchange of bibliography, authority, classification, community information, and holding data in machine-readable format" (Lee, Jeon, & Sung-Kook, 2006a; Bruket, 2008).

Lee, Jeon & Sung-Kook (2006a) proposed MARCXTM, a XTM way to model MARC21 bibliographic elements. Their proposal included both the implementation of the MARC21 format or specification in Topic Maps (through the use of 'a conceptual model' of the MARC

structure) and the representation of real MARC records into it. The last part represented some challenges due to the use of indicators and subfield codes in MARC.

The authors found some limitations of XTM to represent the MARC schema, for instance, the lack of provision "of multiple instances for *occurrences*", the difficulties "to define record schema with [*association types*]" (Lee, Jeon, & Sung-Kook, 2006b). The authors don't extend on the details of these issues.

However, it is commonly agreed (both in the LIS and Topic Maps communities) that MARC has limitations as a format, and the quality of its data is not consistent enough for automatic modelling (see for instance Tennant, 2002). Lee et al. found indeed that "MARC reveals limitations for cataloging bibliographic collections: lack of expandability due to rigorous record formats, difficulties in representing bibliographic relationships, ambiguities in describing MARC records, incompatibilities between other MARC formats, weaknesses in describing bibliographic attributes of digitized resources, and so on" (p.242). The problems of MARC have already unanimously been identified in some parts of the LIS community, and are addressed in the coming standards such as RDA.

The proposal of Lee et al. already described is a representation of a document language at the schema level. However, there are other possibilities for such representation at the record level[19]. This was achieved by Lourdi, Christos & Nikolaidu (2007) (Chapter 3.6.2.1).

Besides, all the FRBRized catalogs with Topic Maps (chapter 3.6.1) were MARC-based, which shows an important conclusion for the representation of document languages with Topic Maps: the need for a conceptual model in between (chapter 4.2).

3.4 Topic Maps and conceptual models

3.4.1 FRBR

Since FRBR is a conceptual model, i.e., an abstract specification of conceptual structures, there is a need for its specific implementation through a data model and a technology. For this purpose, MARC bibliographic records, for instance, are converted into FRBR by creating the

[19] For a differentiation on these types of interoperability see: Chan & Zeng (2006).

association between the entities (there have been several efforts to do this conversion, one of them the algorithm released by OCLC).

This FRBRized MARC is formalized into an ontology, e.g. using Topic Maps or W3C OWL (Aalberg, 2005). This has been the main use of Topic Maps reported in the literature in relation to conceptual models in LIS.

Aalberg (2005) used FRBR as an ontology for a topic map in developing a prototype on music information. This application was intended to serve as a model for the "BIBSYS FRBR project" (in Norway). It implemented the Group 1, 2 and 3, firstly identifying the entities and their relationships, and then extracting the attributes from MARC fields. The main *topic types* were 'Expression', 'Manifestation', 'Person', and 'Work'; the main *association types* were 'Creation', 'Embodiment', and 'Realization'. Aalberg et al. (2006b) present the final report of the project (in Norwegian).

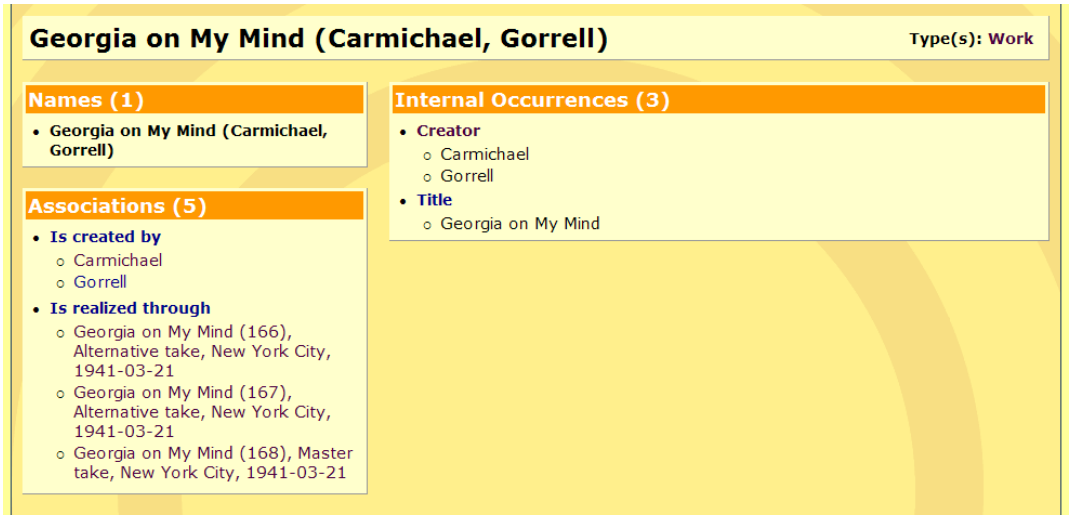

Figure 6: Example of displaying "Work" information in a topic map (Aalberg, 2005?)

Using a records sample from the Grieg collection at the public library of Bergen, Bruket (2008) built a small application to convert MARC records into a FRBR-like model through the use of a topic map. Bruket converted a set of music records in NORMARC (the Norwegian version of MARC) to MARCXML. These records were converted into FRBR using a tool developed by Aalberg et al. (2006a) which uses an XML format (FRBRXML). From this format Bruket proposed two possibilities to create the final XTM file: the use of the Extensible Stylesheet Language Transformation (XSLT) or the creation of an SQL database

which could be exploited through the use of PHP/DOM. The last solution was finally implemented in the application.

Howard (2008) built an ontology for expressing FRBR into a topic map as a proposal to structure the records of the Sound Library of the Australian Broadcasting Corporation (ABC). The following is an example of the display of "Work" information in her topic map:

Figure 7: A musical work conceived as a FRBR work

Prof. Sam Oh was the leader in the development of a prototype proposed to the National Library of Korea to FRBRize its catalog through the use of Topic Maps for the development of the new services of the National Digital Library of Korea, planned to be released in May 2009 (Sam Oh, personal communication, April 12, 2009). Oh (2007, 2008b, 2008c) explains the processes and concepts involved in this application. His solution consisted of using an algorithm (the one created by OCLC) to convert the MARC records to the FRBR model, and a topic map to actually FRBRize the catalog. It appears to be until now the most complete representation and implementation of FRBR in Topic Maps because it includes the three FRBR Groups, their entities, attributes and associations.

These ideas of FRBRizing library catalogs through Topic Maps seem to date back to 2002, as it can be seen in the bibliography listed by Sigel (2004) who reports on the visions of Art Rhyno and the "PHYTEAS project" of Topic Maps as a suitable model to implement the FRBR associations. Sigel (2005) later created a demo topic map to experiment with this idea by using LTM. In his demo, he tried to answer two questions: "how can FRBR be used for the modeling of bibliographic relationships, using Knowledge Technologies like XTM,

OWL/RDF?" and "what are the *Published Subjects* for FRBR to be used as shared vocabularies for XTM and OWL/RDF? Sigel was probably referring to the identifiers that would make possible global information sharing about the FRBR entities, attributes and relationships.

Sigels' demo was used by Vasallo (2006) to make possible the navigation of archive fonds whose descriptions were kept as bibliographies. The MARC-based catalog containing these descriptions was used as a source for the extraction of topics, as well as the documents which were codified in TEI and DALF. For this project a MySql database (containing the MARC records) was turned into XML and then, the topic map was extracted by using a script querying for the database. The "agents" and "fonds" information was extracted from descriptions made in EAD or EAG using XSLT. The relationships were created from the MARC records and from the EAD and EAG files. Vasallo presented this project to the "Archivio di Stato di Pavia" but there is no reported information on its current status.

Finally, Fitch (2002) explains how the principles of RDF and Topic Maps influenced the design of the Australian Literature Gateway (AustLit) project.

3.4.2 CIDOC-CRM

Kivela & Lyytinen (2007a) used information modeled with CIDOC/CRM as data source in creating the Finnish National Gallery and used special tools to model the complex associations to become *associations* in the topic map.

Norrish & Stevenson (2008) structured the relationships between different types of *topics* in the ontology with the CIDOC model, the same was reported by Tuohi, (2005; 2007) and Stevenson, Tuohy & Norrish (2008)

3.5 Applications of Topic Maps in LIS and MLA

The following is a literature review of both actual and potential applications of Topic Maps to MLA and digital libraries in the Humanities, and a summary of the actual applications that claim to be using Topic Maps.

Although Iglesias and Stringer-Hye (2008) called the Topic Maps for Integrated Library Systems (ILS) still "an undelivered promise", partly because there are no vendors currently

using Topic Maps technologies in the creation of those systems, or because it is a relatively new standard (Colmenero, 2005), the literature still reports on various applications in this realm, on some prototypes being developed for experimentation, and on some other promising possibilities.

Topic Maps has been applied for a variety of purposes. These were found mainly in libraries, museums, digital libraries in the Humanities, web publishing, content delivery, and other small applications for specific purposes in the MLA.

The field of **Digital Humanities** has especially benefited from the use of Topic Maps. Bøckman (2006, 2007); Meschini (2005); Walsh (2005) –who presents a literature review on the topic-; Walsh & Dalmau (2006); and Tuohi (2005 and 2008) studied and described the applications of Topic Maps in this field.

These are the details on all those applications:

3.5.1 FRBRization of library catalogs

FRBRization of library catalogs has been one of the most important applications of Topic Maps to library catalogs. Aalberg (2005), Oh (2007, 2008b, 2008c), Bruket (2008) and Howard (2007) show the details of this implementation, which was already described in 3.4.

3.5.2 Integration of library catalogs and records

3.5.2.1 Metadata interoperability

Topic Maps is considered as a model which can be used for mapping different metadata schemas at different levels: Lourdi et al. (2007) concluded this in a study to “propose an efficient method to manage and expose the wealth of complex cultural heritage collections” at the Department of Greek Literature at the University of Athens in Greece. The folklore collection that was digitized and described consisted of the student’s handwritten traveling notebooks, which also included photographs, lyrics and handcrafts collected during the research of the different ways of living in different regions of Greece. Collected from about 1967, they consist of a quite large collection, containing more than 4000 notebooks and 350,000 pages.

This heterogeneity in the resources implied the use of a metadata model that could allow the combination of elements from various metadata standards (descriptive, technical, rights metadata, metadata for educational purposes, and meta-metadata, to indicate the particular standards used for each material type). These metadata also needed to be used at different levels: collection, notebook, chapter, sub-section, page, photograph, objects; and, finally mapped to a common standard –XML based- in order to allow the harvesting of the repository.

In their research, they found that each metadata element belonging to different schemas could be converted into a *topic*, and the *associations* could link the different elements between diverse schemas. These authors affirm that current research in metadata interoperability shows the advantages of using a central ontology that acts as a "hub" or translator from many other different metadata schemas reducing the need for different crosswalks. Topic Maps was suitable to implement this approach (p.211).

In the implementation, a topic map was used to convey two metadata schemas (the source and the target one). The source schema was a (local) collection level schema and the target schema was the Dublin Core Collection Description Application Profile (DC CD AP) which fits as the base for many implementations of the OAI-PMH protocol.

Each metadata element was converted into a topic in the topic map; the topic types were three: descriptive metadata, administrative metadata, and structural metadata. Each *topic* (metadata element) was an instance of any of the three *topic types*. The *associations* were used to relate the correspondent elements or *topics*: each metadata element of the source schema with the semantically related element of the target schema. The association types were defined as: equivalence, refinement (for qualifiers), and hierarchical with the correspondent *role types*.

From this integration, Lourdi et al. concluded that "Topic Maps […] are quite powerful in managing and creating links between different metadata vocabularies and are inherently flexible for defining various kinds of relationships" (p.14).

One interesting possibility –not reported in the literature– derived from this experience would be the integration of Libraries, Archives and Museums at the metadata level (MARC, EAD, CDWA for instance).

The described application was done at the record level. Applications of Topic Maps for interoperability at the schema level consist in the representation of these document languages into Topic Maps. The efforts in this direction were described in 3.3.3.2.

A similar approach was taken in the framework of the Centro di Documentazione Etnografica e di Cultura Apenninica (**CeDECA**) **project**, for a census about cultural heritage in the Oltrepò Pavese. The project used a topic map to integrate different metadata schemas which referred to different kinds entities, bringing perpectives for the possibility of **integrating MLA** through their metadata using a topic map. To achieve it, FRBR was taken into consideration to distinguish the entities, such as agent, objects and access points. The topic map proved to be useful for the integration at three levels: entity level (authority files), structure level (hierarchical relations) and semantic level (subject terms). For the last level "topic maps was used with greater profit". (Vasallo, 2006, p.238).

3.5.2.2 Semantic interoperability

Pharo (2004) presented as a possibility the conversion of subject descriptors into a topic map. The subject descriptors could be represented as *topics*, and the *associations* between *topics* could be made automatically by using, for instance, the frequency of appearance of the same descriptors in different documents. This usage of a topic map in a catalog would allow the users to navigate the conceptual layer created on top of the documents, observing the bibliographic records as *occurrences* of the topics. It would also make possible to "rank documents based on their associative as well as hierarchic relationships. These relationships could be treated as paths with lengths", which could also be used to generate, among others, "usage statistics on path use." (Pharo, 2004, p.11).

Bøckman (2006 and 2007) describes the development of a topic map-based prototype for the Danish National Library of Arts and Architecture. Its main purpose was to allow the browsing of the catalog through the integration of the different and disparate types of materials and vocabularies present in four catalogs and a bibliography. The project focused for that purpose on the integration of the different subject vocabularies in the Humanities, on the specification of the meaning of subject terms based on their context of appearance, and in providing the facilities for browsing and subject access to different types of users: from the novice users who need explicitly stated relationships to know about the domain, to the highly specialized researchers in the Humanities.

One of the main entities of the integration achieved by Bøckman were persons' names, which he found are, together with work names, the most common ones in the Humanities. In the NZETC these entities are the most relevant as well, which led to the development of a system for authority control. Norrish & Stevenson (2008) and Stevenson et al. (2008) describe this application, which was presented in 3.3.2.

Laursen (2006) considered to apply a topic map to integrate diverse subject and library catalogs at the Royal Library in Denmark. A topic map would create a federated catalog of separated systems. The topic map was considered to bring coherence, scalability, and user friendliness. Although the topic map was not finally implemented (there is no information on the reasons), it was considered to have potential to create a unified subject search facility for the many subject catalogues in the library.

In a broader perspective than the application to library catalogs, semantic interoperability was achieved by Farquhar & Bandholdtz (2003) who presented at a NKOS workshop a Topic Maps application to Semantic Network Services. It consisted in the integration of different vocabularies to allow federated searches in different systems. His application appears in the discussion in chapter 4.

At Stanford University Libraries' HighWire Press, Topic Maps is used as a browsing aid based on indexing terms: a topic map based on key words and descriptors gives access to document search by clicking in the *topics* which are navigated through a graphical display.

Integrating different sources, doing mashups is a possibility that Kivela & Lyytinen (2007b) have developed to great extent in the creation of Wandora, a Topic Maps application (designed in the context of museums but expanded to wider possibilities) to combine information from different sources, including third party topic maps. Kivela thinks that Topic Maps is one of the key technologies for knowledge aggregation, which can be combined with Web 2.0 services such as YouTube, Flickr, Delicious, etc (personal communication, April 24, 2009).

3.5.2.3 Linguistic interoperability

Lixin, Jin, & Zhongyi, 2008 present a topic map as a tool for creating a cross-language information retrieval model (Cross Language Information Retrieval -CLIR) for a digital library systems. The article was not possible to review due to language barriers.

3.5.3 Digital libraries, archives and museums

One important application of Topic Maps to digital libraries is in facilitating the navigation of TEI encoded full-text collections.

The New Zealand Electronic Text Centre (NZETC) of the University of Wellington, Australia, was recognized as the most successful application of Topic Maps to digital libraries in cultural domains at the Topic Maps Users Conference in Oslo in 2008. This project makes use of Topic Maps as a tool for presenting TEI-encoded texts due to the limitations of HTML in presenting information that is highly structured (Tuohi, 2007).

Stevenson, Tuohy & Norrish (2008), and Tuohi (2005) describe this implementation in detail: In a structured full-text collection a topic map is used to make explicit the connections between the content of texts, images, and other documents. *Topics* represent authors and publishers, texts, and images, as well as people and places mentioned or depicted in those texts and images. *Occurrences* are used to incorporate relevant external resources which expose structured metadata about their collection. The NZETC digital library currently (2008) contains over 2500 New Zealand and Pacific Island texts ranging from history to literature to language and reference. There are around 110,000 *topics* for people, organisations, places, and texts. A future project includes the expanding of the text mark-up to track linkages such as “allusions” and “influence” between texts, and between ideas expressed in them. In this project, Topic Maps was found to be the best technology to work in conjunction with TEI for structuring, annotating, browsing and navigating very large and diverse digital texts, and a more useful tool than relational databases for metadata management.

Due to the importance of names in this application –a remark also found by Bøckman (2007)–, the NZETC developed in this same framework an application for an entity authority management system. It was described in 3.3.2.3.

John A. Walsh (a researcher on the areas of Digital Humanities and Digital Libraries) and Michele Dalmau (a Usability librarian), both from Indiana University, created in 2005 **The Swinburn project**, a digital collection devoted to the life and work of Victorian poet Algernon Charles Swinburne. “The project aims to unite digitized texts—encoded in XML according to the Text Encoding Initiative Guidelines—and semantic web technologies, such as Topic Maps, to construct a complex database of nineteenth-century British culture with Swinburne at its center” (Dalmau & Walsh, 2007).

Walsh (2005) presents an application derived from this project called **TM4DH** (Topic Maps for Digital Humanities), an open source topic map toolkit for the presentation and navigation of topic maps conforming to the XTM 1.0 specification.

Walsh and dalmau (2006) used in the framework of the Swinburne project a small usability study to compare a Topic Maps-driven interface with an HTML regular interface. Some of the key findings were that the Topic Maps interface invites serendipitious discovery; and despite mediocre task performance and negative satisfaction ratings, participants felt the Topic Map interface had potential, but the interface of Topic Maps was harder to comprehend. The study led them to redesign the interface.

In 2008 Walsh and Dalmau won the "The Herbert White Collaborative Award from Indiana University Libraries" to extend the topic map in order to include more texts and *topics* as well as to apply the outcomes of the Swinburn Project to new online literary collections such as the "Victorian Women Writers Project" and the "Wright American Fiction". Their winning proposal also included a training and guidelines module to document their encoding practices and policies. Michel Dalmau is the editor of the "TEI in Libraries: Guidelines for Best Practices Working Group" and is working on a new edition of the text to be released in November 2009 at the TEI Members' Meeting in Ann Arbor, Michigan.

The **Finnish National Gallery (FNG)**[20] is the largest art museum organization in Finland. It contains three Museums: Ateneum Art Museum, Museum of Contemporary Art Kiasma, and Sinebrychoff Art Museum. In addition to these art museums, FNG contains several cross-organization departments such as the Central Art Archives, the Conservation Department, the Administration and Services, and the Finnish National Gallery Library (Kivela, 2009). The FNG developed a project to integrate a Web service, strongly focusing on their user needs, to their collections management database (Liukkonen, 2008).

Wandora is a general purpose knowledge extraction, management, and publishing application based on Topic Maps and Java[21]. It was developed at Grip Studios Interactive in 2000 with the idea of applying Topic Maps to museum collections in Finland. It is a free and open source tool which has since then been successfully used in Finland for projects at the FNG and in other areas (Kivela & Lyytinen, 2007).

[20] http://www.fng.fi

[21] http://www.wandora.org

In the context of **Archive collections,** image collections are considered to be an area of implementation of Topic Maps. Lewenberger et al. (2006) used a topic map in the "Living Memory" project, a cooperative effort of various institutions in Germany to give access to visual resources in different media that was originated to document a big scale urban planning project. The topic map served both to design a thesaurus specific for the application, which was based on the Getty Art and Architecture Thesaurus, and the navigation and searching tools for the user.

Ramalho, Librelotto & Rangel Henriques (2006) report on "Metamorphosis", a Topic Maps-based tool designed for corporate archives, in the creation of **The Emigration Museum**. It is a digital library on the topic of the Portuguese emigrants to Brazil which tries not only to collect information on each emigrant, but to find the connections between their social influences and activities in their travel destinations.

3.5.4 Subject guides, pathways

Tramullas & Garrido (2006) made a study of Spanish university libraries to develop an application for the creation of **subject portals or pathways** called Potnia which used Topic Maps together with RDF and Dublin Core. "Subject gateways are Internet services which support systematic resource discovery. They provide links to resources (documents, objects, sites or services), predominantly accessible via the Internet." (Koch 2000, as cited in Tramullas & Garrido, 2006, p.1). One of the best examples is the Resource Description Network (RDN) developed in UK (Tramullas & Garrido, 2006, p.1). In this application, Topic Maps was found to be useful to refine and adjust the search processes, as it establishes points of contact between the keywords which were ignored by the traditional process of treatment and by the retrieval of information used initially". (Tramullas & Garrido, 2006, p.5).

One particular library service based on library catalogs is the elaboration of **subject guides**. This has been considered by Iglesias and Stringer-Hye (2008) as "currently the most visible and widespread evidence that topic maps are indeed making inroads into evolving library technologies" (p.17). Peng & Ke (2008) describe how Topic Maps was applied to build the Chung Hua University Library pathfinder (or subject guide), changing its previous simple HTML base to a topic map built with the TM4L tool.

3.5.5 Other applications

The capabilities of Topic Maps in bringing contextual semantic information (Lewenberger, 2006, p.110) proved to be advantageous to new users in the Digital Humanities. Both Bøckman (2007) and the Swinburne Project confirmed this in their applications of Topic Maps. In the Swinburne Project, their usability study showed that "for students, the additional contextual information provided by the glossary and encyclopedic reference features inherent in the Topic Maps metadata standard is critical for understanding obscure and unfamiliar references and allusions in literary texts." (Dalmau & Walsh, 2007). The needs of both expert and novice users in the context of the Humanities (and perhaps in other areas as well) opens up the issue –already noticed by Michel and Dalmau in their Swinburne project– of **Topic Maps as a teaching and research tool**. This thesis doesn't cover that area, but the author of this thesis has found in the course of the research more applications of Topic Maps to Education than to LIS.

For **small scale online applications in MLA** the use of Topic Maps could be suitable. Bøckman (2007) suggested to use Topic Maps to cover specific thematic areas such as exhibitions or educational projects[22]. This has been already done in Hungary, where the National Library uses a topic map for a e-learning application on Hungarian literature for secondary school students. MLA could also take advantage of experiences where Topic Maps has been successfully implemented, such as the online delivery of newspaper archive content (Stevenson & Styron, 2006), or the creation of temporary exhibition websites, such as "The National Treasures", an implementation of Topic Maps for a traveling exhibition of a collection of items from the Australian State and National libraries which toured the country between 2006/2007. The publication of selected media documents produced during a computer fans event Kivela & Lyytinen (2004) and the report on the use of Topic Maps for a digital collection on independently produced movies Ahmed (2007) could give some insights on more potential applications of Topic Maps in MLA.

[22] The original conclusion in Danish says: "Af hensyn til realismen er det desuden afgørende, at emnekortene kan opbygges som led i forskellige selvstændige projekter, der har et specifikt tematisk fokus – fx som understøttelse til udstillinger, projektarbejder, undervisningsforløb eller institutionelle samlinger – og derefter samles og dermed både binde projekt-tematikkerne og biblioteksmaterialerne bedre sammen i en synlig, associativ relationsstruktur." (Bøckman, 2007, p.42)

Finally, Sigel (2006) pointed some of the **Social Sciences** out as possible scenarios for the application of Topic Maps, and Lynne & Miller (2005) reported on the use of Topic Maps for **visualizing searching results** from digital libraries in cultural domains.

3.6 Topic Maps and epistemology

Topic Maps has been presented by the Topic Maps community as a "paradigm shift" or as a "revolution" (Johannesen, 2004; Pepper, Park, J., & Hunting, 2002), and as a "doctrine" (Newcomb & Biesunski, 2003). The authors who claim this explain that it has changed the way to look at organizing information, from a perspective that is traditionally *document-centric* to one that is *subject-centric*.

The *subject-centric view* states that we think in terms of concepts and associations between concepts, instead of in terms of documents and/or applications. Organizing information in a *subject-centric* way entails using the *subjects* as a point of departure for searches and for representations. In this view, even documents are subjects in the sense that they (but not only them) can be the objects of description and representation, that is, "subject of conversation […], a hub around which data resources can orbit" (Newcomb, 2003, p. 43, as cited by Hjørland, 2006a). Hjørland (2006a) summarizes this view: "In TM terminology […] a given document is seen as an occurrence of a topic. In indexing theory what "occurs" is information about a given concept."

Besides this claim, Topic Maps has also received some epistemological and philosophical attention in the literature, where Topic Maps has been considered closely related to the "library ideals" and aims "to preserve and encourage knowledge" (Johannesen, 2008). Johannesen has fervently and consistently reiterated in his writings that Topic Maps is a model close to the everlasting ideals pursued by libraries and that it can help to achieve them on a global scale, due mainly to the fact that it is a standard (for vendors for instance) and to the possibilities it offers for reusing and sharing resources as well as already existing ontologies do. Also some epistemological questions as to what knowledge means and what it means to represent it are pointed out by Johannesen (2006).

Sigel (2003) has also given some insights into the epistemological implications of Topic Maps for KO and KOP: For him, some of the already existing problems in KO recur in Topic Maps. One of them is "the representation of incomplete knowledge" and the tension between

universal and domain specific KOSs. Since KO deals with structuring knowledge, and knowledge comes from interpretation, both epistemology and hermeneutics would be part of subject analysis and representation, as well as the social construction of meaning (p.402), which is an obvious concern for Topic Maps as a standard, a model, and a vision for representing information.

Finally, Bøckman (2006 and 2007) reflects on the logical positivistic view implicit in the idea that *subjects* can be uniquely identified and that there is an "all-encompassing scientific system based on logical inferences[23]" (p.39).

[23] One of the fragments in the original text, in Danish, says: "Prospektet lyder måske fascinerende, men det er ikke desto mindre tale om en gammel, falsk melodi, nemlig en version af den logiske positivismes sprogrensningsstrategi. Der består en betydelig analogi mellem de logiske positivisters idé om et altomfattende videnskabeligt system baseret på logiske slutninger fra entydige empiriske udsagn og emnekort-miljøets forestillinger om et ekspansivt netværk af assertive forbindelser mellem entydigt definerede emner." (Bøkman, 2007, p.38).

4. Synthesis: Topic Maps and Knowledge and Information Organization

This chapter explains the relationship between Topic Maps and LIS by combining the concepts and principles of Knowledge and Information Organization (chapter 2.1 to 2.4) with the concepts and applications of Topic Maps (chapter 2.5 and 3). This provides insights into their conceptual relations, the reasons why the applications are possible, and some future perspectives.

4.1 Topic Maps and bibliographic languages

The components of bibliographic languages – vocabulary, syntax, semantics and pragmatics – will serve as our point of departure for understanding how Topic Maps relates to bibliographic languages.

4.1.1 Vocabulary

The vocabulary of a bibliographic language constitutes its building blocks. It is the central element upon which the other three components (syntax, semantics and pragmatics) have been built in the LIS tradition for the purpose of indexing, classifying and describing documents, retrieving information and allowing communication among members of specific discourse communities. Bibliographic languages, as special constructs, try to reduce the lexical ambiguity present in natural languages, which is the cause of failures in precision and recall, and misunderstanding in communication.

Figure 8 shows the most common elements of the vocabularies in bibliographic languages.

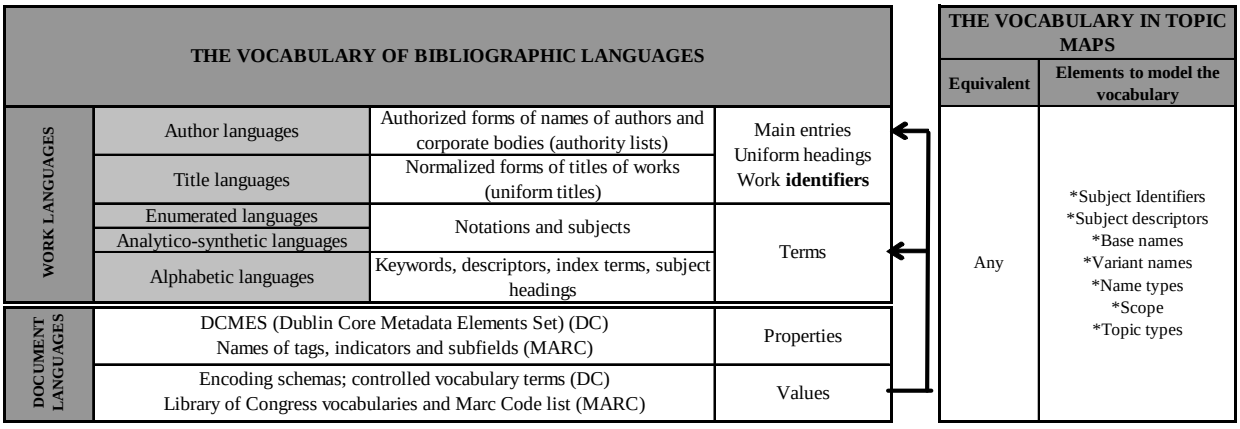

Figure 8: Vocabulary elements

The problem of lexical ambiguity is of a different nature in work languages and in document languages.

In work languages, the main problems are synonymy (multiple terms for a single concept), homonymy (a single term refers to multiples concepts), and polysemy (a single term refers to multiple concepts that have closely related meanings). The goal of work languages is that "each term refers to only one concept and that every concept is designated by only one term" (Svenonius p.89).

In document languages, the main problems are, first, to choose an appropriate set of properties to describe documents (which vary according to the needs of different domains and communities), and second, to establish rules for assigning values to those properties.

We could refer to this process as to a 'normalization of names', which is the purpose of the whole apparatus built by bibliographic languages: "vocabulary control is the *sine qua non* of information organization. Information is not organized if it is scattered or if its collocation is cluttered." (Svenonius, 2000, p.89). Collocation and integration are then achieved in bibliographic languages through vocabulary control based on the forms of names.

Topic Maps, on the other hand, doesn't specify any vocabulary; that task is left to the topic map author: "*occurrence types* – like *association types* and *role types* – may be freely defined … to suit the needs of users and applications" (Pepper, 2008b). Garshol (2008) also pointed to this characteristic saying that the use of "open vocabularies" in ontologies allow their creators to "define the terms as they go along, adding and changing terms to better fit the

dataset, and they do not have to follow any guidelines as to what terms can or cannot be used" (p.386). "Topic Maps is such a general way that needs you to fill in the blanks and the context (to a certain degree)". Johannesen (2006).

The reason for this openness and flexibility is that Topic Maps is not a bibliographic language (section 4.2 will cover this in more detail). But then, how can Topic Maps solve the problems of lexical ambiguity that bibliographic languages try to solve? Their differences would have to be found, not in their vocabulary, but in the techniques they respectively provide to solve the problem of identification of entities.

4.1.1.1 Identification

Identity in bibliographic languages (as it can be seen in Figure 7) is based on the creation of 'authorized forms' of names. The form of the name becomes the referent point for non-authorized or related terms, that is, for all its variations. The mechanism is simple: "choosing one name as authoritative and mapping alternate names to it frequently is all that is needed for vocabulary control" (Svenonius, 2000, p.91). Enormous amounts of time and effort have been spent in creating rules and standards to "select the preferred form" of a name: the ISO and NISO standards for thesauri, the ISBD, the AACR2, and currently, RDA.

Topic Maps is based on the idea that what can solve the problems of lexical ambiguity is identification, not through names but through *subject indicators*. This is discussed next.

4.1.1.2 Problems with "names" and identification

The Topic Maps perspective of the vocabulary of bibliographic languages is that it is term-centric, that is, based on the "nominal attributes" (Svenonius, p.87) of entities, in the terms or 'signifiers' of the concepts. On the contrary, Topic Maps claims to be 'concept-centric', i.e. focused on the 'signified' things.

However this is arguable, because if we think about how to make common use of the same identifiers for the purpose of interoperability, or how to attach all the *variant names* to the identified subject, we can see the traits of an old problem that has its roots in meaning and human communications.

Bibliographic languages are in principle, as Topic Maps, 'concept-centric'. In subject languages, for instance, "an indexing term is 'the representation of a concept'" (Aitchison, p.17), just as in Topic Maps a *topic* is a proxy for its *subject*. Both bibliographic languages and Topic Maps seek identity based on concepts and their meanings.

The nature of the problem of identity is of a similar nature to both of them: one term per concept ≈ one topic per subject ≈ one URI per 'proxy'). The problem seems to be then not *how* to say that these are identical (a term sting or a URI) but to know *when* they are identical. Using both URIs or "nominal attributes" to establish identity, the problem would become related to meaning. Bøckman (2007) presents an interesting discussion of some of these problems in Topic Maps from a linguistic point of view. But what would the 'answer' be in a Topic Maps perspective to the problem of knowing when two *subjects* should have the same *subject identifier*? It would be: through the use of *subject indicators*.

Subject indicators are meant to aid the negotiation of meaning (when they are shared in the form of *PSIs*). They are intended to be read by humans (as opposed to the *subject identifiers* which are meant to be interpreted by computers). *Subject indicators* give humans evidence of the meaning that will allow them to "unambiguously identify the *subject* represented by a *topic*" (ISO/IEC 13250-2). It is thus *subject indicators* and not *subject identifiers* that are the core of the identification mechanism provided by Topic Maps.

From the bibliographic languages perspective, this identification principle doesn't seem to be new. For instance, scope notes in thesauri, and definition notes and the other five types of notes in DDC are intended to serve the same purposes (Batley, 2005, p.35), i.e., to indicate information about the identity of concepts to allow the user of the bibliographic language to select the appropriate term (or, in Topic Maps terms, to aid the topic map author in selecting an appropriate *subject identifier* for a topic).

However – and this could be applied further on – the big difference between Topic Maps and bibliographic languages is that Topic Maps provides a standardized model and interchange syntax for addressing the issue of identity in a digital environment and on a global scale. This sole feature of having a scope note (or *subject indicator* – which is in principle the same thing), represents inherent advantages for bibliographic languages, because the forced economy imposed by what Steve Newcomb called the technology of paper (Newcomb & Biezunski, 2003) reduced the subject languages to be used simply as tools for description and

classification, and not as sources of information for the user in their own right (the cataloger or the end user). For instance, in a paper based thesaurus, the scope notes, the relations between the terms, and all the other notes, remained in the "tool" when the right terms were picked up by the cataloguer. This is what Peter Morville defined as an 'indexing thesaurus' (Morville, 2007, p.211). Lately, with digital technologies, these printed thesauri could migrate in the form of hyperlinked terms, becoming 'searching thesauri' (Morville, 2007, p.211), bringing the thesaurus as a whole to the browsing and searching mechanisms (see chapter 4.4.5.1). Topic Maps is based on XML, which adds to the advantages of Topic Maps over searching thesauri which are based on HTML. However, the main difference between scope notes and *subject indicators* is that the later ones are, together with the *subject identifier* part of a single mechanism for identification in Topic Maps, and will be represented accordingly in XTM.

Vocabularies and systems nowadays are meant to work and to be shared globally, not in the confines of a single application or in the boundaries of centralized and hierarchical forms of distribution that have as a consequence the passive uses of 'authorized' forms. The so-called 'term-based' mechanism of bibliographic languages is, as Norrish & Stevenson (2008) observed it, basically limited and inadequate for our times, because, "a single name or form of a name is unnecessarily privileged [and] indeed, stands in as the entity itself)" (p.2).

Today and always language –and names– constantly change. Besides the current tendencies in digital information ("the third order" as David Weinberger called it) and in self organization (the idea that 'Here comes everybody' pictured by Clay Shirky) make the name-based selection of "identity" an obsolete practice.

But, as it was said before, *subject identifiers* are the basis for merging. How then could this be achieved? How can one achieve interoperability if there is not a common *subject identifier*? These are issues that go beyond the scope of the Topic Maps family of standards and transcend to the sphere of the Topic Maps community, where different points of view and approaches to similar problems are constantly discussed and brought to the level of different implementations and initiatives. One of those approaches will be presented when discussing the issue of global identity (section 5.2.1).

To summarize, we could state (with Steve Pepper, personal communication, May 22, 2009) that the principles that could represent the Topic Maps approach to the problem of names and identification are:

1. Subjects have many names.
2. Users should be allowed to use any name that they prefer
3. Collocation cannot be based on names
4. Collocation has to be based on identifiers
5. Identifiers must be global

However, besides Topic Maps, other solutions also present mechanisms for avoiding the problem of identification through "nominal attributes": relational databases for instance, or RDF. The comparison with these is outside the scope of this thesis, but the directions to explore their differences would be – and this could be applied further on as well – in "the distinction between direct and indirect identification, which is supported in both the TMDM and the interchange syntaxes", which has resounded in the recognition of Topic Maps as a 'subject-centric' approach and the basis upon which to characterize the "strengths and weaknesses of alternative approaches" to Topic Maps (Pepper, 2008b).

Finally, this last issue leads us to comment on what the **identification of documents** could mean according to the Topic Maps principles and elements. This is a rich area for further research. For the purpose of this synthesis, we can state that in Topic Maps, when a document (in the LIS perspective) is an object of description, it is regarded as a *subject*, i.e. as any other entity in real or possible worlds. Its identification then, is subject to the same principles mentioned above. For instance, a document can have many names (different titles), and its identification would be given by a *subject identifier* – now, how does the LIS community identify its objects of description? That is another area for further exploration, but some hints will be presented later.

Regarding Hjørland's equation of *occurrences* to 'information resources' (see 3.2), we can see the different approaches to 'information resources' from a 'document languages' perspective and from a Topic Maps perspective: An *occurrence* is actually a *relationship* between an information resource and a *topic*, and the information resource can be either an externally stored document or a string (or some other data value) stored internally in the topic map. In a topic map, all *occurrences* are considered as information resources, while in a

bibliographic perspective, only *external occurrences* would be considered 'documents' (an exception made in the case of, for example, an *internal occurrence* which contains a full text).

In a bibliographic perspective, the identification of 'documents' (the 'physical embodiments' refered by Svenonius) with Topic Maps would require then the use of *subject locators* (a possible URL), and *subject indicators*, which could potentially include a bibliographic record (or metadata) as a form of unambiguous identification. The *subject identifier* could correspond then to a DOI URI, for instance.

To complement the description of 'bibliographic documents', the *occurrence type* can equates to a 'document type' (Hjørland, 2006).

The main issues concerning the identification of documents in a bibliographic perspective (and in a semantic Web perspective as well) are related to the problems of direct identification and indirect identification, to the use of *subject locators* as *subject identifiers*. Once more, this is an area of research on its own.

4.1.2 Syntax[24]

Due to the artificiality of bibliographic languages, the terms need to be constructed, and some guidance, in the form of a system of rules, indicates how to structure them. This is what is called the syntax of the bibliographic languages. Examples of syntactic rules (given by the pragmatics of the bibliographic languages) are those that indicate the order of the elements of a name and last name, or those in a physical description, or the use of punctuation marks in the use of qualifiers or notations, or the rules to construct compound names in a thesaurus. The traditional cataloging rules used in LIS provide rules and guidelines for such constructions. The AACR2, ISBD, ISAAR(CPF), the ALA Filing Rules, the Library of Congress Filing Rules and the ISO standard for thesaurus construction are a few examples of them.

Figure 9 shows the main ways in which syntax is specified in bibliographic languages.

[24] It is important to notice, that the meaning of the term "syntax" in the *scope* of 'bibliographic languages' (section 2.2.3.3.) is different than the one in the *scope* of the Topic Maps community: the interchange format of the topic maps, i.e. XTM.

THE SYNTAX OF BIBLIOGRAPHIC LANGUAGES			THE SYNTAX IN TOPIC MAPS	
			Equivalent	Elements to model the vocabulary
WORK LANGUAGES	Author languages	Formation of statements and strings Disambiguation.	None	(Rules of thumb, TMCL)
	Title languages	Formation of statements and strings Disambiguation.		
	Enumerated languages	Citation order (depends on the schema)		
	Analytico-synthetic languages	Citation order, notational synthesis (ex.the formula PMEST), Phase relationships (optional). Rules for postcoordination		
	Alphabetic languages	Formation of strings, statements, chains and subdivisions. Compound terms rules Disambiguation.		
DOCUMENT LANGUAGES	DCSV (Dublin Core Structured Values) Guidelines for creation of content (AACR2 and ISBD for derived and assigned terms)			

Figure 9: The syntax of bibliographic languages

On the contrary, Topic Maps neither specifies a vocabulary nor 'syntax' in the sense used by Svenonius. Some "rules of thumb" exist to fill the need for agreeing on the forms that names should assume and in choosing base names, basically for the purpose of sorting and displaying names. Besides many other features related to the construction of the ontology, the Topic Map Constraint Language (TMCL) (ISO/IEC CD 18048) formalizes some of these aspects by specifying which data types and forms of names are valid. TMCL rules can for example state that "Topics of type person must have two explicit names, the full name and a nickname," but they don't go to the level of detail of specifying the exact form of the nickname or the full name. Even though the need for complex syntaxes is reduced with the use of Topic Maps elements, this is one of the areas where Topic Maps can benefit from KO and bibliographic languages expertise and tradition.

Besides, Topic Maps could serve as potential experimentations in current research on how to use the existing syntax of bibliographic languages to create facets from subject headings or UDC notations. Due to the fact that the syntax of bibliographic languages is simple (Svenonius, 132) it can be automated by few algorithms. There are projects already in progress, using different models, where the capabilities of Topic Maps could be put in practice: the project to use the subdivisions in LCSH as facets, and the different facets in UDC or DDC to display subject information in library catalogs: for example in the German

National library's projects to use SKOS and connecting to Linked Data –a project of the W3C Consortium for sharing URIs in RDF applications– in their MelvilSearch and CrissCross projects.

4.1.3 Semantics

4.1.3.1 Referential semantics

Referential semantics relates to the techniques used to limit the meanings or referents of terms trying to solve the problems of homonymy and polysemy. An example could be the word *reading* which refers both to the process and to the city.

The basic mechanism that bibliographic languages have created to achieve the goal of designating every object or concept by only one term has been the creation of 'disambiguation' techniques. Svenonius (2000, p.148) explains some of the methods of semantic disambiguation in use (domain specification, qualifiers, notes and hierarchy).

According to Garshol (2004), disambiguation in Topic Maps "is not necessary, and the types, occurrences, and associations of the topics will generally distinguish them anyway." (p.385).

The following is a possible Topic Maps view of the mechanisms used for disambiguation in bibliographic languages:

1. Domain specification

Svenonius (2000, p.148) explains that the domains of discourse that create the different bibliographic languages could serve as a source to limit their possible referents. That is, if two bibliographic languages belonging to two different domains are merged and two terms are the same, their concepts could be deduced according to their domain of provenance, e.g. *reading* (if one educational thesaurus and a geographical one are merged). However, Svenonius also points to the difficulties that this represents in general or universal subject languages, and says that the best practice is to qualify words with multiple referents.

Observing Garshol's (2004) example of the use of *scope* in Topic Maps, it could be possible to think that this feature of Topic Maps could be used to solve the domain specification problem:

> "For example, different terminologies are often used within a single organization, whether because of regional differences, or because of differences in corporate cultures within the organization. Topic maps can support this by scoping the names with topics representing the corporate cultures or areas where the terms are used. This allows the topic map to say that 'corporate culture A calls this region "APAC", while culture B calls it "Asia-Pacific"'. Users can then in their profile (which may be permanent, or just for a single visit to the site) state their preferred terminology, and the site can then display the correct names for each topic for them." (p.385)

However, *scope* is basically used to solve the synonym problem (to delimit the context of validity of an assertion: multiple names for single subjects) while the 'domain specification' used in bibliographic languages solves basically the problem of homonymy and polysemy (multi-referential words). The problem of homonymy and polysemy are not solved then both in bibliographic languages and Topic Maps through the use of qualifiers, but they are of different nature.

2. Qualifiers

In bibliographic languages this could include, for example, the GMDs (general material designators). In Topic Maps this function could be covered by *topic types*, *associations* and *occurrences*. For example: in the case when two *topics* have the same name (for example 'Paris'), the most common way to disambiguate (as natural languages do), would be its category (*topic type*): 'Paris (city)', 'Paris (god)'. If it is the case that there are two cities with the name 'Paris', a third step in disambiguating would be the *association type* 'located in': Paris (city-France), Paris (city-United States). If there are two cities in France with the name 'Paris' an *occurrence type* could be use as a third disambiguator, and so on. However, the differences in the mechanism used by Topic Maps and bibliographic languages by providing qualifiers is that in bibliographic languages they are part of the name string (as it is for example in the names of author names which have an added date of birth as qualifier or in the case of the GMD). In Topic Maps, the different blocks correspond to different elements of the model (*topic type, association type, occurrences*) and can be automatically displayed when they correspond to the same string in the base name. *Scope* is used as well, as in the example showed above, to add specifications of the context of validity of the assertions when the three mentioned elements are not enough.

3. Notes

Bibliographic languages (mainly subject languages) make use of notes (for example scope notes in alphabetical languages, or definition notes in classification languages) to explain the use of a term in a certain context. They can be treated in Topic Maps through *occurrences* or through *subject indicators*, as it was described before.

4. Hierarchy

In bibliographic languages the contextualization of the meaning of terms can be deduced by the context given by the relations to other terms (for instance, in the Dumb-down principle in DC). This is the best example of the use of relational semantics to providing meaning. In Topic Maps, associations are used to provide semantic information. One of the main advantages in this aspect is that a topic map is a "searching KOS" (4.1.1.2) and will always display the associations as part of the information for a given *topic*. Besides, the non-hierarchical nature of Topic Maps allows making explicit all the types of hierarchical relations (see 4.4) given, thus providing more semantic information on a given topic.

In summary, the disambiguation problem presented by bibliographic languages is observed by Topic Maps in two directions:

1. For computers: the homonymy problem is solved by the used of *PSIs*
2. For the users: The disambiguation is a problem of display, and it is solved through the use of *topic types*, *association types*, *occurrences* (in the case of homonymy and polysemy), and by *scope* (in the case of synonymy).

4.1.3.2 Category semantics

In Svenonius' terms, category semantics refers to the part of the vocabulary of the bibliographic language that specifies its categories or facets: "facets are the grammatical categories of a bibliographic language" (p.57). The advantages of such categories are, according to Svenonius, firstly, that they can be used to formulate syntax rules and decide on access points; and secondly, that they allow managing the vocabulary and building the relationships that will connect the terms.

THE CATEGORIAL SEMANTICS OF BIBLIOGRAPHIC LANGUAGES			THE CATEGORIAL SEMANTICS IN TOPIC MAPS	
			Equivalent	Elements to model the vocabulary
WORK LANGUAGES	Author languages	Personal names Corporate names	None	Topic types Name types
	Title languages	Uniform titles / Series titles (Assigned titles: Title proper, Paralell title, Variant title, Abreviated title, Devised Title)		
	Enumerated languages	Classes		Topics or Topic types Association types
	Analytico-synthetic languages	Classes, facets, subfacets (arrays), foci		
	Alphabetic languages	Topic – Place – Time – Form (in LCSH) Top Term (TT) in thesauri Types of relations		
DOCUMENT LANGUAGES	The properties		None	Association types Occurrence types Name types

Figure 10: The categorial semantics of bibliographic languages

This categorization function would often correspond in Topic Maps to the use of *topic types*. Garshol (2004), however, says that *topic types* are an extra capability of Topic Maps and not found in any bibliographic languages: "[…] in traditional techniques the set of terms is in a sense flat, since there is no way to distinguish different kinds of terms" (p.386), Figure 10 shows, on the contrary, that bibliographic languages do use category semantics. The problems in using them as *typing topics* when representing bibliographic languages with Topic Maps comes from the differences in the types of hierarchies (e.g. subject classification hierarchies, type-hierarchies, etc.), as it will be described in 4.1.3.3. and which may or may not be considered as *typing topics*.

The main implication of category semantics for representing bibliographic languages, especially in work languages, with Topic Maps is that it could eventually indicate the types of things existing in those vocabularies, that is, one part of the ontology of a topic map (the *typing topics*) or, in other cases, the types of relational structures to be modeled through *association types*. Yi (2008) observed that the existence of taxonomies already in place in the existing digital libraries would make it easier to convert them to "machine understandable" structures through the use of ontologies, rather than developing new structures for that purpose (p.?). Section 4.2. expands on this idea, and tries to explain the relation of conceptual models to the category semantics.

This kind of semantics is not clearly identified in document languages. In the MARC bibliographic format, for instance, it could be perhaps said that the main tags give categorial

information on the entities they contain; e.g., 200- Title and Title-Related Fields indicates that the kinds of entities it describes are ‘titles’. But this is scarce and difficult to identify. Those who are modeled MARC with FRBR have noticed these limitations on the expressivity of this language. Dublin Core (DC) is more clear in specifying ‘Classes’, e.g. “agent” (“for person, organization, and software agent”), “bibliographic resource” (“for book, article, or other documentary resource”).

4.1.3.3 Relational semantics

One of the main characteristics that make Topic Maps be considered an evolution of existing KOS is that it offers a model suitable to express any type of relational semantics of bibliographic languages. This is possible mainly because of the already explained “nominal based” system used by bibliographic languages to identify concepts; i.e., this mechanism has imposed upon them the use of limited types of associations between terms. For instance, a thesaurus makes use of associations by limiting them to five types: BT, NT, RT, USE, UF.

As described in 3.3. some efforts have been made within the Topic Maps community to represent bibliographic languages in topic maps: for instance, to represent thesauri (Ahmed, 2003), to represent faceted classification (Garhsol, 2004), and to represent hierarchies. At the same time, as said before, Topic Maps has the ability to express the relational semantics of KOS. Also, Topic Maps has been said to be able to represent synonym rings and taxonomies (Garshol, 2004). At the same time, Figure 11, representing the “increasing structural complexity among controlled vocabularies”, is well known in the LIS community.

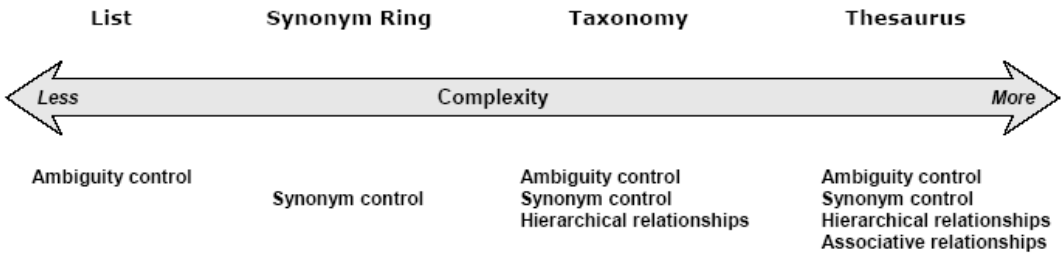

Figure 11: The complexity scale (ANSI/NISO Z39.19-2005, p.17)

These confusing terminologies can obscure the meaning of what is being represented in respect to the relational semantics of bibliographic languages. Figure 12 is an attempt to

distinguish them, based on Svenonius' categorization of the relationships among terms (equivalence, related-term and hierarchical):

BIBLIOGRAPHIC LANGUAGES		TYPE OF BIBLIOGRAPHIC LANGUAGES	RELATIONAL SEMANTICS	RELATIONAL STRUCTURE
WORK LANGUAGES	Author languages	Authority files	Equivalence	Term list
	Title languages	Title Authority Headings	Equivalence	Term list
	Enumerated classification languages	Enumerated classification schemas	(Equivalence)	Taxonomy
			Hierarchical	
	Analytico-synthetic languages	Faceted classification schemas	Hierarchical	(Faceted structure)
			(Equivalence)	
			(Near-relatedness)	
		Faceted thesauri	Hierarchical	Thesaurus facet
			Equivalence	
			Near-relatedness	
	Alphabetic languages	Subject authority lists	Equivalence	Synonym ring
		Thesauri	Equivalence	Thesaurus
			Hierarchical	
			Near-relatedness	
DOCUMENT LANGUAGES	(metadata schemas)		(Hierarchical -instantial- relationships between properties and values)	(At the level of properties and values)

Figure 12: Relational semantics and relational structures

As it can be seen, only alphabetic languages and, to certain extent, analytico-synthetic languages, make use of the three different types of relational semantics.

Basically, the three types of relational semantics in bibliographic languages can be expressed in Topic Maps with the use of *association types* and *topic types*. *Topics* and *topic names* and the relational structures can be expressed through a certain combination of those.

Following there is a brief description of the three types of relational semantics in the perspectives of bibliographic languages and Topic Maps:

Equivalence relationships

Equivalence relationships address the problem of ambiguity in names by relating terms that are synonyms. The basic techniques to do this in bibliographic languages, as explained by Svenonius (2000, p.158) are the use of "cross-references", which can relate one term to another through "Use" indications or the use of "double (or multiple) indexing", making

terms point to the same authorized term. The last technique is used in classification (same classification number) and book indexes (same page) (p.158).

As explained before, the Topic Maps mechanisms for identification remove the need for preferred forms of names, and all the terms are related to each other without statements about which is preferred above the others. In this sense: "every topic is really a synonym ring" (Garshol, 2004, p.385). Synonym rings can be represented in Topic Maps by *topics* and *topic names*. However, in the case that preferred terms have to be kept for specific purposes (for indicating classification numbers for example) –which doesn't occur in a synonym ring–, the different terms can be indicated through *scope* and the preferred terms can be used as the *default names* in an *unconstrained scope*. The *scoping topics* can indicate the reasons why those terms are not preferred. This is usually lost in subject languages where non-descriptors simply keep the reference to the preferred terms without further explanations for the preference (as in the case of indexing thesauri as opposed to searching thesauri). Sometimes an indication that a term is obsolete or has been modified, deleted or replaced is included in an historical note that may or not be displayed to the users. By using these features of Topic Maps, all term lists can be represented.

Hierarchical relationships

A hierarchy is an arrangement of concepts in a tree structure based on subordinate-superordinate (or parent-child) relationships. According to Broughton (2007?) there are three main kinds of hierarchical relationships: 'thing-kind' or 'genus-species' (sometimes called taxonomic relationship), 'whole-part', and 'instantive or instantial'. These relationships are also often called semantic or paradigmatic because they are inherent to what constitutes the concept. In bibliographic languages, the main hierarchical relationships, according to Svenonius, are genus-species and a type not mentioned by Broughton, perspective hierarchies.

Svenonius states that "the genus-species relationship, also known as the inclusion relationship, is the classic hierarchical relationship with the properties of reflexivity, transitivity, and anti-symmetry. It has another property as well, which in the computer literature is called inheritance and in the classification theory hierarchical force, whereby what is true of a given class (Furniture) is true of all classes it subsumes (Chairs, Tables, and so on)" (p.164).

The perspective hierarchies, on the other hand, as explained by Svenonius (2000) are those that indicate the class of a concept. In her example, the term 'insect' can be described from

the point of view of "agricultural pests, disease carriers, food, art representation, and control technology" (p.151). These relationships are not considered hierarchical in Topic Maps, but of near-relatedness instead, and can be described by *association types* of the form 'belongs to class'.

Despite being an associative (as opposed to a hierarchical) model, Topic Maps does have a built-in hierarchical association type, called superclass-subclass (or supertype-subtype; the terms are used interchangeably in Topic Maps). Associations of this type are binary and each of the two role players is by definition a *topic type*. This corresponds exactly to the genus-species relationship.[25]

In general, the responsibility for defining *association types* is left to the topic map author. This means that any kind of hierarchical relationship can be represented, including those referred to by Broughton as 'instantial' and 'whole-part', and others, such as containment hierarchies ("located in"), etc.

The hierarchical relationships (mainly the taxonomical one) are most useful in connection with enumerative classification languages, and have been modeled with Topic Maps (see section 3.3.2.2 –classification languages).

Göktürk, Rosvall & Gutierrez (2004) studied the 'perspective hierarchies' and their modeling with Topic Maps, highlighting the quality of Topic Maps of "being neutral towards hierarchies". Their research has implications for the use of Topic Maps and perspective relationships "for building knowledge bases by collaboration of multiple experts" (p.?). Pharo (2004) also noticed that traditional KOSs like thesauri are hierarchically organized, but that Topic Maps inspires to create more non-hierachic/network-like systems. Pharo also describes in this a challenge for the implementation of non-hierarchical associations, for instance in websites based on thesauri which are implicitly hierarchical:

> "In my opinion it is particularly interesting to analyse the idea of topic maps in light of thesaurus construction and facetted classification. The topic map-environment is particularly clear in stating that topic maps open up for non-hierarchical knowledge organisation systems. This, however, necessitates the construction of non-hierarchical

[25] There is also another built-in association type, called class-instance (or type-instance), which relates a topic to its type. Associations of this type do *not* form hierarchical structures.

> ontologies. In practice most, if not all, web sites organised using topic maps will consists of one or more hierarchies with cross-hierarchy references. Rather than dismissing hierarchies as the structural format of the ontology it is more important to discuss how to implement the non-hierarchical references. A thesaurus is an example of a hierarchical ontology that permits explicit references between" (Pharo, 2004)

Consonant with this idea, Luckeneder et al. (2001) also found that hierarchies have been found to be a "proper organization" for "structuring well defined fields of knowledge" while not efficient enough for fields that are unstructured (p.746). This would be another reason to explain why Topic Maps have also been found to be suitable for those unstructured domains (see 4.3.1).

Document languages are not structured to the point of introducing these kinds of hierarchical relations. Therefore, there has been a need to create conceptual models that can explain the underlying relationship structures not only between records, but also between entities and their attributes. In the FRBR model for example, whole/part and part-to-part relationships are specified in order to allow the discrimination and description of containers and carriers, and components and aggregates (Tillet, 2003). This type of hierarchy also describes the realizations of works in expressions, the embodiments of expressions in manifestations, and the exemplifications of manifestations in items.

The basic structure used to express hierarchical relationships is a 'taxonomy'. Garshol (2004) uses this term "to mean a subject-based classification that arranges the terms in the controlled vocabulary into a hierarchy without doing anything further" (p.381). The main bibliographic languages that make use of taxonomies are the enumerative classification schemas and the analytico-synthetic ones. How they may possibly be modeled with Topic Maps is explained in 4.2. Author and title languages use few hierarchical relationships. The most obvious case would be to show subordinate-superordinate relations among names (Svenonius, 2000, p.57). Indexing languages such as thesauri only make use of two "properties": BT/NT to name hierarchical relationships, which can include many different types not explicitly expressed.

As it could be seen in the literature review (chapter 3.3.2.2), modeling bibliographic languages with Topic Maps implies the recognition of the different types of hierarchical relational semantics present in them. Ahmed (2003) for instance, insisted on the fact that "in order to really make use of a faceted classification system it is important to be able to

determine what facets are defined in the classification system, and the type or types of association which define the hierarchical relationship between classes in each facet." This means, to extend the vocabularies to name the different types of relational semantics (chapter 4.4.2).

Related-term relationships

Known also as "non-hierarchical" relations, this type of association includes every possible association that is neither equivalent nor hierarchical. They are named as RT in thesauri and as 'see' and 'see also' notes in classification languages: "traditional guidelines for constructing subject languages define related-term relationships vaguely, negatively, and broadly to include all semantic relationships, except equivalence and hierarchy" (Svenonius, 2000, p.160).

This lack of expressivity in thesauri and the like has been the main reason, as said before, to consider Topic Maps as an "evolution" of bibliographic languages, because Topic Maps enables the different kinds of associations expressed under the single type "near-relatedness" to be specified explicitly. Associative relationships appear mainly in thesauri and in faceted classification schemas, in the form of related terms (RT). The inability to specify the nature of such relationships more explicitly is one of the main arguments in favour of modeling bibliographic languages with Topic Maps, because its open vocabulary would allow making explicit the many types of associations that are grouped into a single RT. Exactly how this can be done will be described in 4.2.

4.1.4 Pragmatics

This component of the bibliographic languages refers to their use or application (Svenonius, p.58). Cataloguing rules within the LIS domain address this element, but they mainly cover document languages. Among the subject languages, Svenonius only recognizes the DDC as containing a certain level of pragmatics. For that reason, we will focus here on the Knowledge Organization Processes (KOP) in KO: cataloging, classification and indexing.

One of the possible impacts of Topic Maps in KOP comes from its subject-centric view. In traditional KOP (i.e. cataloguing, classification and indexing), the center is the document, "the item at hand" which is described through the use of a set of "properties" that come from the controlled vocabularies (work languages, derived terms) as well as from the item itself

(assigned terms). Would the subject-centric view of Topic Maps imply a kind of inverse process? From starting cataloging based on the "item at hand" (as recommended by the AACR2) to starting with the 'subject at hand'? Would this imply a complete redesign of the LIS traditional bibliographic systems? Would the LIS community move from transferring bibliographic records to 'transfer' topic maps?

This scenario could be one of the many possible futures of cataloging. However, in an immediate perspective, what Topic Maps has to offer to traditional KOP is an enhancement of bibliographic systems themselves, by making possible (as it was seen in 3.5.1 and 3.5.2.) the application of conceptual models and metadata and semantic interoperability between disparate systems. As with any other technology, the human and intellectual effort is (still) needed. But the associativeness of the Topic Maps model is closer to the relational way of thinking that documentalists and librarians (in different perspectives though) have had through the years.

Considering the 'item at hand' not as a physical embodiment to be described but as a subject to be enriched with meaningful associations would allow librarians to think in terms of "works", which can trigger a set not only of procedures but attitudes towards cataloging, classifying and indexing.

4.2 Topic Maps as a bibliographic meta-language

As it has been observed in the previous sections, the main relation of Topic Maps with Knowledge and Information Organization seems to be that it can represent and extend all the knowledge representations used in LIS. The following sections tries to explore how this fact could be explained and concludes with the main implications of having this potential.

The representation capabilities of Topic Maps in a LIS perspective, could be summarized in four areas that go in progression (in parenthesis some observations which would correspond to a Topic Map view):

- In a basic perspective, documents could be considered as representations of human knowledge through for example linguistic or audiovisual expressions. (And as representation of concepts and notions through *subjects*)
- Work languages are representations of the knowledge registered in documents. (*Topics* and topic maps are representations of the *subjects* registered in

documents plus concepts or things existing in real or fictional worlds that haven't been registered)

- Document languages are representations of documents and work languages, i.e. their 'bibliographic descriptions'[26]
- Conceptual models are representations of the conceptual structure of records, in this case in the cultural heritage realm. (In a Topic Map view this stage is related to the identification of ontological elements for the topic map, according to the category semantics)

Topic Maps, as it has been seen, can represent both 'documents', 'document languages', 'work languages', and 'conceptual models': the representation of 'documents' was possible to observe in initiatives that link Topic Maps and TEI for the 'mapping' of full texts. (3.5.3). The representation of 'document languages' was seen in 3.3.3. Also the representation of 'work languages' (3.3.2) and 'conceptual models' (3.4.).

This could lead us to define Topic Maps from a Knowledge and Information Organization perspective as a **bibliographic meta-language**. Auillans et al. (2002) have actually defined Topic Maps as "a meta-language for structuring meta-data" (p.70). But if meta-data is considered as a "bibliographic description", then this consideration is partial, because Topic Maps 'structures' many other things as it has been shown.

The reasons that make this possible have been described in 4.1 and can be summarized saying that Topic Maps has the capabilities of:

- representing the 'vocabulary' of the bibliographic languages by extending them through the use of a more flexible identification mechanism fully adapted to digital environments and the Web,
- representing both the relational structures and their implicit relational semantics. Also the extension of the latter one through the use of vocabularies to make explicit the non-identified relationships of bibliographic languages,
- providing mechanisms to carry out the functions of the syntax in bibliographic languages; and

[26] See chapter 2.2.2.1 for an explanation on how this concept was interpreted from Svenonius. There is a discussion further on.

- representing both the relational semantics and relational structures implicit in the categorial semantics of bibliographic languages (which is related to the 'ontology' of a domain).

The most obvious conclusion for these facts is expressed by Garshol (2004), he referring to classification schemes:

> "A summary of the relationship between topic maps and traditional classification schemes might be that topic maps are not so much an extension of the traditional schemes as on a higher level. That is, thesauri extend taxonomies, by adding more built-in relationships and properties. Topic maps do not add to a fixed vocabulary, but provide a more flexible model with an open vocabulary. A consequence of this is that topic maps can actually represent taxonomies, thesauri, faceted classification, synonym rings, and authority files, ***simply by using the fixed vocabularies of these classifications as a topic map vocabulary***." (Garshol, 2004, my emphasis)

Using this abstract level view (which at a first glance may appear simplistic compared to the complexities of a MARC format, or the DDC, just to mention two) we could try to see how this representation of bibliographic languages is possible, and why Topic Maps is a bibliographic meta-language:

As I explained in 2.2.2.1, the differences then between document languages and work languages are not in the entities they describe, but in how they describe them. For instance, Pepper (2008a) states that "assigning metadata to resources is equivalent to making statements about *topics*." (Pepper, 2008a). In this sense a description of the type property-value would correspond to a document language that makes statements about any kind of 'entity', while a work language would serve as the attributes for such a description. However, work languages, to serve as attributes, have to make a "description" of its entities as well, but not in the form of a property but of a semantic arrangement. In a work language, the property-values also exist, as in BT: 'term'. But this is just the label for an internal semantics (the relational) that is the one that gives sense to the construction. As Garshol (2004) explains: "Metadata only relates objects to subjects", it doesn't arrange them (p.381). In this perspective, any language that is used to describe properties is a document language. Any entity can become a value, and any attribute can become an entity. For example, the 'document language' that describes a thesaurus, is: BT, NT, UF/USE, RT, Scope note

The entities of a work language can be concepts, but also names (of people, as in author languages; or of documents, as in title languages). In this sense, the rationale to model both types of bibliographic languages with Topic Maps would be:

- Modeling document languages with Topic Maps[27]
    - Does the property have naming semantics? ----model it with *Name*
    - Does the property have identification? ----model it with *Identifier*
    - Is the value of this property something which is a *subject* in its own right or something the **users of the system** want to regard as a *subject*? ----model it with a *Topic*
    - If not (the previous one) ----model it with an *occurrence.*
- Modeling work languages with Topic Maps

The rationale for this representation, as I could observe, is the identification of the relational structure and the relational semantics of the bibliographic language. The best practice found in the literature is to follow 'design patterns' for representing work languages. This could be (Figure 11):

- For author and title languages: Modeling the term list structure
- For subject authority lists: Modeling the synonym ring structure
- For enumerated classification languages: Modeling the taxonomical structure
- For thesaurus: Modeling the thesauri structure (which has been already done, as seen in 3.3.2.2)
- For faceted classification schemas: Modeling the faceted structure (which has been already done, as seen in 3.3.2.2)
- For faceted thesauri: Modeling the thesaurus facet structure
- What follows after modeling the underlying relational structure is an interpretation of the different types of relational semantics: if they are hierarchical, equivalent or near relatedness:
    - If they are hierarchical: the different types of hierarchies would define if using the inherent relational structures of Topic Maps or *association types* that name the type of association..

[27] (Steve Pepper, personal communication, May 29, 2009)

- If they are equivalent: Use *Names*, *variant names* and *name types*
- If they are of near relatedness: use *association types* for expressing the underlying relations among the terms: the extending naming capabilities of Topic Maps (as explained in 4.3.2.)

A general principle that applies for modeling with Topic Maps is "to represent the values of properties as topics wherever possible, and this means using *associations* rather than *occurrences* to represent property/value pairs." (Steve Pepper, personal communication, May 29, 2009). The more this principle is followed, the more associative the topic map is.

This view at a high level of abstraction (representing entities, relations and attributes, underlying relational structures and "reusing vocabularies") poses the question of what are the implications and consequences of having a"bibliographic meta-language", of what would be the purpose of adding those vocabularies to a different model.

The obvious answers such as migration to digital environments, reuse of misused elements (such as scope notes and other types of notes and relations) appear as big advantages, but they are nothing compared with the main benefit promised by Topic Maps: semantic integration:

4.3 Topic Maps as a common meta-language

Once documents, document languages, work languages and conceptual frameworks are represented with Topic Maps through the rationale already showed, they can be integrated: *topic types, topics, association types, associations, occurrences types, internal and external occurrences, names, variant names, identifiers, subject indicators* will represent all the kinds of things that exist in Information and Knowledge Organization (i.e., it has the potential of becoming the 'ontology' of the bibliographic realm, in this case in the sense given to it by Sowa, 2000). All becoming part, in principle, of a single sharable topic map.

The differences then between documents, document languages, work languages and conceptual frameworks built with the 'document' as a center are regarded from an upper level as a single structure where everything (a property, a value, a name, a note, a subject entry, etc.) is possible to represent as a *topic* and become an 'object' of description in its own.

Combining all these languages is like combining a traditional library catalog with an encyclopedia, where almost each single property referred in the record is an access point to

three different kinds of thing: information on the property, connection with the same entities that share the attribute, and to the information resources (both as entire documents –papers, books, articles- or as fragments within those documents). The 'encyclopedia' needs though a conceptual structure, the 'ontology', that can indicate the types of things to be modelled. In bibliographic languages this role corresponds to the categorial semantics.

Topic Maps is known as a language for expressing ontologies. Garshol (2004) says that they are "created to be an ontology framework for information retrieval" (p.378), and Pharo (2008) for instance, defines it as a standard for representing ontologies. However, as it has been explained in 2.6, there are different meanings of 'ontology'.

In this case of Topic Maps as a bibliographic meta-language, it refers to the 'kinds of things' present in the bibliographic realm, i.e. the entities, attributes and relations that are defined in the conceptual frameworks. That is why, the categorial semantics is the necessary bridge for any representation of a bibliographic language with Topic Maps, and that's why the conceptual models play the vital role of providing the 'vocabularies' for the types of entities, attributes and relationships of the bibliographic languages built with a semantic and logical structure.

This meaning of 'ontology' is broader than the one in 'terminological ontologies'. However, in both cases, the representational leader in Figure 4 should have Topic Maps on top, according to what has been described.

Finally, the fact that Topic Maps is at a higher level of abstraction, makes it independent of any specific technology and then, in principle, able to be transmitted, reused and shared across the space and along time.

Besides, one additional characteristic of Topic Maps as a bibliographic meta-language is that it is an ISO standard. However, an area that is out of the scope of this thesis, is how Topic Maps could interoperate with existing standards, for instance:

- ISO/TC 37/SC4
- ANSI/NISO Z39.19-2005
- ISO 5127, Information and documentation — Vocabulary
- ISO 21127:2006
- The Vocabulary Markup Language (Voc-ML)

- eXchangeable Faceted Metadata Language – XFML Core
- SKOS – Simple Knowledge Organization System

Moreover, its 'syntax', XTM could be studied as well in comparison with MADS, MODS, MARCXML and simmilar XML languages.

Besides, the main problematic area for research would be how to represent the specificities of each special KOS, that is, the different ways in the abstract structures have been implemented in particular cases, how to 'deduce' and identify the different types of semantics present for example in DDC, UDC, The Agrovoc thesaurus, and the like).

4.4 Enhancing Knowledge and Information Organization

The first potential application of Topic Maps to enhance bibliographic languages is within their **migration to digital environments and to the Web**, by using the current standards and syntaxes to be machine-understandable. Colmenero (2005) found this adaptation to be one of the main uses of Topic Maps (p.78). For instance, tools like "Tema Tres"[28] developed in Argentina for the creation of thesauri, includes as one of its functionalities the exportation to an XTM file.

Other possibilities are described:

4.4.1 Naming and vocabulary building

Topic Maps provides the task of naming and "mapping names" with a flexibility that goes beyond the possibilities offered by bibliographic languages and their equivalence relationships. This capability, in conjunction with other Topic Maps capabilities, allows not only the indication of the *variant* or alternate names, but an explanation of why these forms are variants and of which type (e.g. "XML Topic Maps" is a variant of "topic maps" because it is a "misnomer" of the type "obsolete term"). This is done through a combination of naming capabilities and *scope* (which will be discussed later). Garshol explains the advantages of this mechanism:

> "This enables the creator of the topic map to define a kind of language for describing names. For example, one could create a topic for topic maps and give it the name 'topic

[28] www.r020.com.ar/tematres/

maps' (with empty scope), but also the name 'topic navigation maps', in the scope 'obsolete'. This would have the same effect as the USE/UF construct in thesauri, or as an authority file, but is more powerful, since one can say *why* the term 'topic navigation maps' should not be used. (The reason being, of course, that it is obsolete.) (Garshol, 2004)"

Another capability of Topic Maps related to the flexibility in naming is the use of *parameters* for variant names, which are sets of topics that can give indications to machines of when to select the most appropriate *variant* for a certain task.

Finally, basing identity in *subject*s and not in "nominal attributes" has one more consequence: every name can become an "access point" which allows more fluid navigation and use of non-authoritative forms: this opens up possibilities for the combination of highly structured views of organizing information (usually represented by MLA), and the tendencies of bottom-up approaches in building vocabularies. For instance, Folch, Lahlou, & Benoît (2000) created a topic map to act as a "semantic interface to the document collection" allowing navigation through the different semantic classes that were created both from a pre-defined list of categories and from the emerging topics in the corpus" (p.272). For these authors, when text collections are heterogeneous and do not correspond to circumscribed domains, and when they are dynamic and "loosely structured" (as in the case of digital information), top down classifications based on pre-defined categories become insufficient. In these cases, automated classification, based on an inductive approach, proved to be useful (Folch, Lahlou, & Benoît, 2000).

User participation in so-called social tagging could be an additional use of *variant names*. The combination of identification mechanisms that are at the same time structured but flexible (for instance, the use of an *identifier* but with the possibility to attach any number of names to it without hierarchies or preferences) makes Topic Maps a model that favors less hierarchical views biased to preferred forms of terms. This factor can also contribute to the consideration of Topic Maps as a model that is in between the highly structured visions of structuring information in the Web and the vision of the Social Web or Sociosemantic Web, that is, to combine bottom up with top-down Information Organization perspectives (expressed in the form of folksonomies). The project Fuzzzy.com[29] is an attempt to do this.

[29] www.fuzzzy.com

A consequence of this is that Topic Maps seems to be more suitable to be applied in domains with blurred or not fixed terminologies. For example Yi (2008), found that "Topic Maps is […] more appropriate for nonscientific fields than RDF/OWL because Topic Maps can represent multiple meanings for each term and can build complex relationships among terms" (p.?). Bøckman (2006) also found that in the humanistic domain the most predominant feature is the presence of multiple variants of names, and in that sense, Topic Maps could support the browsing function among them while respecting the need for interpretation by the user. Bøckman (2007) showed how Topic Maps proved to be useful in providing context information that could aid in the construction of the meanings of concepts, "which is in line with the hermeneutic approach of the humanities" (p.?). In this sense, Topic Maps seems to provide a flexible model that allows working with the structuring of information in domains where there is no established terminology. It facilitates dealing with the problem that Svenonius (2000) called "lexical indeterminacy", i.e. when a subject can be named but not unambiguously defined.

It seems that Topic Maps plays a better role in representing knowledge for specific domains, getting closer to the users' natural language (Colmenero 2005, p.?). Some of these capabilities of Topic Maps have already been put to practice, as it was seen in 3.5.

Sigel (2003) concludes that "knowledge structures as represented in Topic Maps should be sources and results of shared understanding processes. Since understanding is open-ended, it is a requirement that we can also represent incomplete or partial knowledge." (p.399). This could lead to research into possible applications of Topic Maps in the process of "controlled vocabulary building" to observe how Topic Maps capabilities could aid in the "negotiation of meaning" in a domain with non fixed terminology.

4.4.2 Extending bibliographic languages

Until now, it has been observed that the representation capability of Topic Maps of bibliographic languages covers the flexibility in naming due to the use of the identification mechanisms, the limitation of the referential semantics. It also covers their relational semantics, which can be not only exactly modeled with Topic Maps, but enhanced and extended as well. However, as explained in 3.3. this capability of Topic Maps (to represent bibliographic languages, and specially work languages) is focused on their capacity to extend

relationships, which is cited as the main reason to consider them an "evolution" of existing KOS.

This is because the LIS community had already noticed the need for this extension. Svenonius (2000) for example states that "traditional guidelines for constructing subject languages define related-term relationships vaguely, negatively, and broadly to include all semantic relationships, except equivalence and hierarchy", and express clearly that "the breaking down of generic related-term relationships into groups of more specific relationships would seem to be inevitable in the general evolution of subject languages toward specificity and formalism." (p.160,162). Hjørland (2006) also commented on this necessity of all bibliographic languages to be enriched by term definitions, notes on term usage and the more defined relationships.

The Medical Sciences, for example, is one of the domains where explicit near-relatedness associations are an urgent need (Hjørland, 2006). One application where Topic Maps was used for this purpose in this domain is explained in Okada et al. (2007), who used a topic map to display the relations of the words used by medical departments.

Besides near-relatedness associations, hierarchical ones also claim for more explicit statements: "Most seriously, perhaps, if the different types of hierarchical relationships are not distinguished, any attempt to translate among several classifications or simply to achieve compatibility in retrieval is impaired" (Svenonius, p.165). Bøckman (2007) also found that the vocabularies in the Humanities are fully populated with persons' names and work names, and those ones are not usually compatible with thesauri (which keep subject terms) in the sense that there is no meaningful way to map them. He presents one example[30] stating that it will be a challenge to integrate for example the proper name 'Abildgaard' into a thesaurus structure that specifies the NT or BT for period names such as 'classicism' and 'romance' under the BT 'modern times' which has as the same time 'historical time' as BT, since 'Abildgaard' cannot have NT or BT. In Topic Maps this could be solved easily with the use of *topic types* for 'person' and an *association type* named for example 'belonged to historical

[30] The original text is in Danish: "Det er en selvstændig udfordring, som det ses nedenfor, at skulle integrere disse to hovedtyper af emneord i en samlet organisering, primært fordi navne-emneordenes evne til at indgå i fx taksonomiske relationer (BT/NT) er omvendt proportional med deres præcision (periodebetegnelser som 'klassicisme' og 'romantik' kan have 'nyere tid' som BT, som igen kan have 'historisk tid' som BT, men 'Abildgaard' kan hverken have NT eller BT)." (Bøckma n, 2007, p.37)

period'; or, with the CIDOC terminology, using an association type 'was born' which will be linked to artistic and cultural periods in history.

Lewenberger et al. (2006) applied Topic Maps in modeling the Getty Art and Architecture Thesaurus, by dissolving its hierarchical composition into association types that could express the implicit semantic relations.

Equivalent relationships also suffer from this limitation of bibliographic languages in describing relations:

> "See references are sometimes used to link antonyms, on the grounds that antonyms represent opposite points on a continuum scale and, thus, really refer to the same concept. However (again), the use of one device for multiple purposes has the potential to cause trouble. In this particular case it has the potential to cause serious miscommunication in retrieval, to deteriorate precision, and to obstruct transparent linking." (Svenonius 2000, p.159)

In Topic Maps, antonyms would be considered as different *topics*, related by the association "opposite of". Besides, through the use of *scope*, it is possible to explain why a term is non-preferred.

This capability of Topic Maps shows the promising possibilities for the improvement of bibliographic languages. Indeed, thesauri (and ontologies –'terminological ontologies') have been considered to be enhanced by using explicit relationships through Topic Maps. Yi (2008) insisted on the need for further developing the capacity of Topic Maps in allowing the expansion of thesauri by making explicit the associative-relatedness relationships. Garshol (2004, p. 386) and Kongsbakk (2004) also describe this idea. Besides, the effect of this "expansion" that Topic Maps make of thesauri has already proved to be useful in retrieval systems by having a positive influence on "improv[ing] recall and making the search time shorter for relationship-based queries than for those of a thesaurus-based information retrieval (TIR) system […] The results of this study attest to the potential of Topic Maps-based ontology to improve information retrieval system performance through better support for associative relationships between terms belonging to different hierarchies by providing explicit relationships among resources" (Yi, 2008).

Nevertheless, extending bibliographic languages in this sense would appear at a first sight to be costly and difficult to implement: “mapping various types of related term relationships from one controlled vocabulary to another can lead to massive intellectual effort to resolve subtle differences between them” (ANSI/NISO Z39.19-2005). Pharo (2004) has also manifested the need for research on the mechanisms to implement the associations that are not hierarchical in subject languages.

However, even though the implementation would require additional efforts, these are reduced by the existence of vocabularies that can provide normalized terms for the *association types* and *role types* that that could be reused (Sigel, 2006, slide 55). Few examples are:

- There are several languages providing *association types*: Perreault (1965, as cited in Sigel, 2006) published a classified list of 120 relationships, including proposals from various classification specialists such as Ranganathan; and Schmitz-Esser (1999, as cited in Sigel, 2006) compiled a controlled vocabulary on *association types*, which was used in the Integrative Cross Language Ontology (ICLO). The CIDOC/CRM which is inherently defined to provide “definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.”
- Relator codes that already exist could be reused as *role types*. “The value of this ‘role’ information becomes very apparent in light of FRBR. We need to regain the lost link of relator terms and codes in our bibliographic records” (Tillet, 2003).
- ‘Document types’ could be reused as *occurrence roles*

Chen and Chen (2001) referred to these enhancements as being a “relationship-centric approach [which] should be highlighted as a key framework for knowledge organization theory”. Topic Maps provides the mechanisms for implementing it.

4.4.3 Merging and integration

Sigel (2000) sees that “semantic interoperability” is one of the main problems of KO, and that even if any solution has proved yet to be “convincing” enough, Topic Maps could present many advantages because it is a “tool with which knowledge structures could be maintained more easily, and thus more time could be dedicated towards better quality” (p.9).

Merging is one of the mechanisms that Topic Maps presents for the integration of different semantic structures (bibliographic languages for instance). Merging is in the origins of Topic Maps, which idea was to merge the different back of book indexes. What makes *merging* possible is the establishment of common identification for *topics* through the creation of *Published Subject Identifiers* and *Published Subject Descriptors* for global (or shared in particular domains) identification.

Pepper (2008) defines *merging* as the single most powerful feature of Topic Maps. It could serve, in his words, for global aggregation of knowledge, for information integration across repositories, for sharing and reusing taxonomies, for automating content aggregation, for distributed knowledge management, and global knowledge federation (Pepper, 2008).

The experiences described in 3.5.2 don't report explicitly to have benefited from the use of *merging* as the standard defines the mechanism. It is possible that the integration they achieved was done through the creation of a single topic map that served as a central point for modeling and integrating the separate systems, but not of different topic maps than lately were merged. *Merging* is also used into a single topic map, but this functionality was not reported in the literature.

Johannesen & Pearce (2004) state that "one data model is easy to share due to Topic Maps built-in ability to merge." And that merging allows reusing data from many projects: "It hence becomes easy to create applications that are grander in scale, reusing from other projects as needed. You don't have to wait for those projects to create import and export facilities." (Johannesen & Pearce, 2004)

The idea, although powerful in principle, has the limitations given by the establishment of identity, not limited by the mechanisms of Topic Maps but by the semantics of language and the problems of human (or machine) interpretation.

Sigel (2000) synthesize this: "with Topic Maps one can express conceptual structures, but, of course, Topic Maps do not come up with valid fusion strategies. Without some background in KO or comparable experience, naive merging will result in a big pile of rubbish in which all context will be lost" (p.?).

Merging as aggregating is a simplistic view of meaning. However, this problem is not inherent to Topic Maps, and if the agreements on conceptual models are in place, merging would represent a promising capability for the integration of library catalogs.

For instance, Sigel (2006) sees that one of the potentials of Topic Maps for KO is the aggregation of knowledge from semantically heterogeneous sources. This was used by Farquhar & Bandholdtz (2003) who applied Topic Maps to create a semantic network service to "harmonize environmental information from national and state authorities" in Germany. The topic map allowed the integration of different information resources through the integration of different vocabularies using a topic map; those controlled vocabularies were an environmental thesaurus, a national gazetteer and an environmental chronology. The context of the use of Topic Maps in a 'semantic network', as an enabler of the semantic interoperability in the system architecture can be seen in the following graphic:

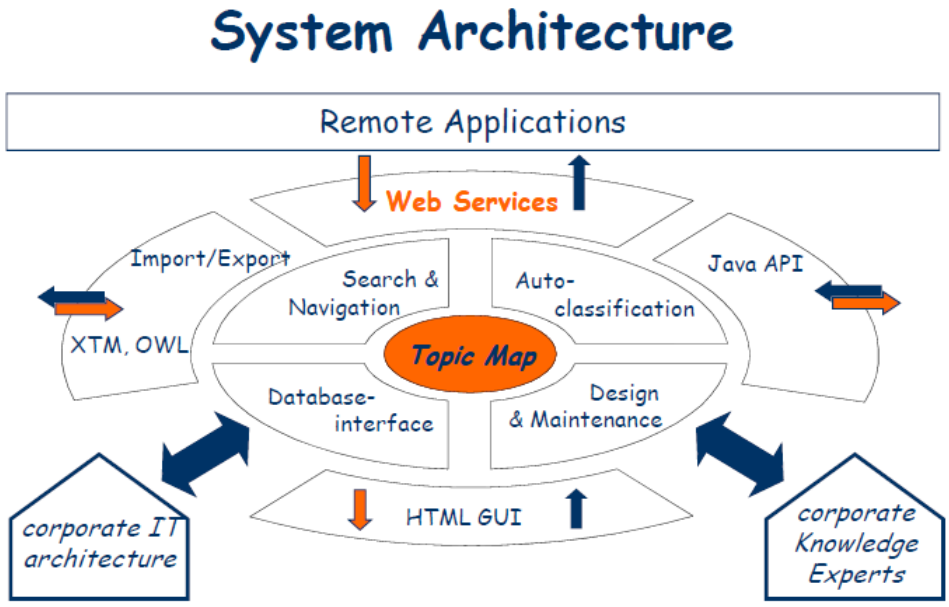

Figure 13: A Topic Maps system architecture
Farquhar & Bandholdtz (2003)

Johannesen (2008) also considers Topic Maps as one medium to integrate different thesauri and other classification schemes. This creation of associations between different metadata schemas to allow interoperability was actually achieved and explained by Lourdi, Christos & Nikolaidu (2007) who concluded in their study that "developing mappings with Topic Maps mostly facilitates the semantic interoperability of the metadata schemas, while OAI-PMH focuses on syntactic issues" (p.210).

Merging is not without problems. Bockman (2007) for example relates this issue with the epistemological problems of uniquely identifying subjects, saying that even merging two topic maps on the same topic the result may not be necessarily meaningful.

4.4.4 Context and user perspective

Scope is an element of the Topic Maps model specifically designed to address those problems of variations due to differences in contexts and perspectives. Svenonius, who didn't know about Topic Maps when she wrote her book, expressed this need as follows:

> "A practical problem in exploiting perspective information for disambiguation purposes is how to package it for users. Imagine a user, as a first step in an online search, keying in the work Freedom. Ideally, the system would behave like a good reference librarian and respond by asking the user "What kind of Freedom do you mean –freedom of speech, Freedom of the press, Freedom of religion?" it would elicit perspective information form a universal classification and present it to users in a friendly manner. The possibility of using perspective information in this way has long been understood but has been slow to be realized in practice. There are difficulties. One experiment which sought to exploit perspective information in retrieval, produced the disconcerting finding that users did not find it helpful. But this may have been due not to the information but to its manner of presentation. In any case, work needs to be done to harness the disambiguating power of traditional classifications for use in online retrieval." (Svenonius, p.153)

The experiences reported to date, however, only show applications of *scope* in the context of corporate cultures in an organization and in multilingual systems. In the first case, Garshol (2004) explains that *scope* would allow users to see in their profiles their preferred terminologies from a set of vocabularies coming from different organizations or bodies. In the second case, Lixin and Zhongyi (2008) followed the same approach in creating a cross-language information retrieval model using Topic Maps. Garshol (2004) also points to the possibilities of *scope* for multilingualism in applications:

> "Since topics representing languages can also be used as scopes [t]hus we could give the topic for topic maps (English) the name 'emne kart' (Norwegian) and have the system display the names in the user's preferred language. Since scopes consist of *topics*, it is up

to the creator of the topic map to define their own language for describing the *names* of *subjects*.')" (p.?).

The most interesting implication of *scope* however, is its use in combination with the identification mechanisms of Topic Maps. To the question of 'what happens if there is no agreement in different domains in the *subject descriptor?*' one could think on the possibilities of expressing variations in conceptions and definitions according to authors, tendencies or movements, as well as in their changes in time: "Topic Maps can be employed to express both contradictory discourse community views and subject metadata for knowledge repositories" (Sigel, 2003, p.386)

One of the building blocks of the Scriptorium project (Folch, Lahlou, & Benoît, 2000) was the use of *scope*. It was used to express "the limit of validity of a topic" within a certain context. In this project, the context was defined by the different groups of "semantic classes" (defined as topics) that were produced according to the different text sub-corpus, from which different chunks were formed. Within each of them, the topic characteristics (name, occurrences and type) were limited by the scope feature. Vasallo also (2006) explains as well the use of *variant names* to allow users to use natural language in their searches in an integrated catalog to favor "research success". The experience also proved to help "user education" in the sense that users can be brought, through the topic map, on the "correct" ways of spelling and naming.

Scope is not mandatory though, and names without scope are considered to have unlimited validity for the users of a specific topic map.

According to its founders, Topic Maps doesn't assume a naïve view of the *subject uniqueness* principle, but a possibility of expanding possibilities for expressing the views of conceptions in different domains when these touch a certain common point. (Newcomb & Biezunski, 2003).

4.4.5 Topic Maps and the bibliographic objectives

Following there is a synthesis of the capabilities of Topic Maps to aid the accomplishment of the bibliographic objectives that have impacted information systems such as digital libraries:

4.4.5.1 Finding

In a LIS perspective, “finding” is mainly conceived as an activity to locate items of which some criteria are known. This criteria have traditionally corresponds to what is called an “access point”, which is constructed based on the normalized forms of names. In current catalogs, searches are also performed in the full text of the bibliographic records in an ILS.

Traditionally, “finding” presupposes that certain information (metadata) is known about works, editions or items that are searched. In *topic maps*, since any *topic* can be an “access point”, there is no privileged metadata, and any *topic* can be considered as such.

In Topic Maps, and according to Pepper (2008a), the findability problem is produced because much existing metadata consists of simply string values. Following the principle of modeling *subjects* using *associations* instead of *occurrences*, the navigability increases and thus, the browsing experience. Yi (2008), for instance, showed how collocation produces an effective and efficient searching (p.1898).

At the NZETC, for instance, the finding capability brought by a topic map is described as “accidental knowledge discovery” and as a support for exploration (Tuhoy, 2008). This is an illustrative example from that application:

> “The National Library of New Zealand hosts a full text archive of the Transactions and Proceedings of the Royal Society containing New Zealand science writing. By linking people topics in the NZETC collection to articles authored in the Royal Society collection it is possible to discern an interesting overlap between the 19th century community of New Zealand Pakeha artists and early colonial geologists and botanists” (Stevenson, Tuohy & Norrish (2006, p.2)

For specific searches, that is, by known items, Topic Maps provides a query language, called TMQL, which can facilitate the formulation of complex queries based on the richness expressivity of the *associations*. Portal Federation with TMRAP (Topic Maps Remote Access Protocol) is another possibility for findability with Topic Maps, as well as the full text search.

Other options found in the literature that show how Topic Maps aids the findability objective are: the experiences on access to codified texts, which allow the Integration of metadata and content.

Besides, as it was explained before, Topic Maps allows all knowledge representations in LIS to become searching KOSs instead as indexing KOS. Searching thesauri, for instance, represent enormous improvement for searching and browsing:

> "In other words, your thesaurus can become a true portal, providing a new way to navigate and gain access to a potentially enormous volume of content. A major advantage of the searching thesaurus is that its development and maintenance costs are essentially independent of the volume of the content. On the other hand, it does put much greater demands on the quality of equivalence and mapping" (Morville, 2007, p.212).

The Topic Maps view of "finding" represents a middle point between highly structured models and the full text approach. This has led to associate Topic Maps with the idea of a more semantic web for example (see 5.2.3 for a discussion).

4.4.5.2 Collocating

In a Topic Maps perspective, collocation means to present all the information related to a topic in a single location (web page for instance in the context of digital information).

In Topic Maps, collocating is the principal means by which finding is achieved. According to Newcomb (2006), the most basic thing about Topic Maps is that it allows having "one subject per virtual 'location'". Collocation has to do with "how to aggregate statements (assertions) about the same subject from different resources, and […] the potential of providing new (semantic) knowledge services and products on top." (Sigel, 2006). Yi (2008) showed how the collocation of information can be improved by adding semantic relationships through topic maps (and in general, through the use of ontologies- (p.1898).

One of the creators of Topic Maps used to call *topics* as "binding points". *Topics* are basically a "collocation points". Identity is the mechanism that Topic Maps provide for collocation, as it is explained in 4.1.1.1). In principle, since Topic Maps is a structured way to represent information as opposed to the algorithmic key word search proposed for non ontology based information retrieval systems, it is meant to help achieving collocation and precision. As explained before, bibliographic languages accomplish the collocation objective through vocabulary control (Svenonius, 2000, p.88).

Sigel (2006 and 2003) refers to this objective in Topic Maps as SLUO "Subject Location Uniqueness Objective", i.e. the semantic integration that consist in making accessible from one single location everything that is known about a given *topic*

4.4.5.3 Navigation

Navigation is at the heart of Information Architecture, says Peter Morville. In the origin of the Topic Maps standard there is a goal to enhance navigation. Its original name was "topic navigation maps". The fact that in Topic Maps everything is or can be a *topic,* and that each *topic* is an "access point" makes possible to integrate concepts and descriptions providing facilities for a non hierarchical browsing.

Referential semantics is one of the keys to enable navigation: Garshol (2004, p.388) and Yi (2008) agree in the fact that navigation is aided through an enriched relational semantics (making explicit the associations between concepts).

The navigation of bibliographic records is explained in 3.6. where one of the main conclusions is represented by Pharo (2004) and Bøckman (2006, 2007) who describe how the use of a topic map in catalogs would allow the users to navigate bibliographic records through the integration of different vocabularies and/or catalogs.

Oh (2008a) demonstrated as well how a topic map-based system enhanced its navigation possibilities: among other advantages it required users to use fewer clicks and less browsing, the information structure was more detailed and specific, well structured and related, the users could find all the relevant and related information in one page, and serendipitous findings were well-supported.

Topic Maps was also used as a way to allow the navigation through topics or subjects of structured full-text materials in a digital collection (Stevenson, Tuohy & Norrish, 2008). Peter Morville's ideas of "findability (wayfinding, navigation and retrieval)" and "structured serendipity (the value of unsought finding)" were possible to achieve through the use of Topic Maps. The navigation goal was achieved due to the capability of Topic Maps to allow making explicit the implicit connections and cross references between books and papers. The topic map "provided a meaningful structure based on real world entities through which the user can explore the rich content in the collection and find other relevant online resources" (Stevenson, Tuohy & Norrish, 2006).

Svenonius (2000) recognizes this as one essential objective of bibliographic systems, and says that it “calls for creating all useful associative trails among documents” (Svenonius, p.23). In this sense, the navigation objective would be accomplished if integration is achieved: integration among bibliographic languages, and among bibliographic systems. As it was shown in 4.2, Topic Maps, as a common meta-language makes possible this integration and thus, the navigation.

In conclusion, this is the objective best served by Topic Maps to bibliographic systems, as a result of the collocation objective. Both navigation and collocation facilitate finding objective, which is also complemented by query possibilities.

Navigation is the result of structuring relational knowledge, and to the fact Topic Maps creates what Bowers and Delcambre (2000) called a “superimposed information” structure “which is a layer (the superimposed layer) of data placed over existing information sources (the base layer) to select, combine, highlight, supplement, and provide additional links among selected information elements within the underlying sources”. (p.?)

4.4.5.4 Choice

This objective serves the purpose of distinguishing between two or more entities with similar characteristics Also called “discrimination”. Distinguishing among similar resources is nowadays one of the main concern in the creation of conceptual models for library catalogs. To allow users to discriminate and select, the differences between works, expressions, manifestations and items have to be implemented in catalogs for the clear discrimination of documents.

The use of *topic types* in topic maps would make possible the creation of faceted structures which allow filtering according to diverse criteria. This is the tendency of ILS nowadays.

4.4.5.5 Acquisition

This objective comes from the idea that bibliographic systems (specially document languages) serve the purpose of locating items. However, acquisition can be both digital or physical. The use of *occurrences* and *subject locators* provides this acquisition capability.

Due to the *merging* capabilities of Topic Maps, federated searches of bibliographic records would make possible acquisition at a broader scale.

5. Discussion

5.1 Methodological discussion

The consequences of doing research on a concept from a broad and comprehensive perspective within a limited time frame present obvious methodological limitations: conceptualization, data selection, data analysis, and verification are the most important ones that may have impact in the findings of this thesis.

For instance, applications of Topic Maps to Information Architecture, a flourishing discipline tightly related to LIS, needed to be left apart. The same happened with the uses of Topic Maps in education and research, which is also an area difficultly separable from LIS, especially in the context of academic and school libraries.

Besides, LIS as a discipline lacks agreed-upon definitions on some of its existing and emergent concepts, and it touches many other disciplines and transcends to global and political realms. On the other hand, the Topic Maps community is still in formation and not consolidated enough to present consolidated views on some important issues. Its only official documents are the Standards, and those, besides its Data Model (TMDM), are closer to the Computer Science domain and thus not always possible to grasp for the author of this thesis. Because of this, keeping the scope focused on Topic Maps within LIS was a difficult task to achieve and represented a difficult limitation that will be reflected in this work.

Moreover, there are more methodological implications in presenting a 'state of the art' on the placement of Topic Maps in LIS: for instance, the data selection was limited by the lack of documentation or dissemination of the implementation experiences as well as in their visibility. The sources were of different natures (such as blogs, mailing posts and conference presentations), which presented incomplete or non-verifiable information. Some of the most representative sources were in languages that the author of this thesis couldn't have potential access to, as Danish, Norwegian and Chinese; and some in German as well that couldn't be explored at all due also to time limitations.

Regarding the data analysis, as mentioned in section 1.3.5, I used in this thesis a software for qualitative analysis, called Atlas.ti. The main purpose was to codify the texts and audio materials (recordings from conferences and lectures), to identify topics of interest and create

families of codes which were meant to serve to create the categories for analysis. Besides, to a certain extent, it was used to write memos for observations on quotations and codes that were used in the initial steps of the analysis. Modeling concepts and relations with a topic map would have served, in principle, for the same purpose that was served by Atlas.ti, that's why my initial methodological design included the creation of a topic map for data analysis. However, due to my limitations with no previous experience in using a topic map for GT, I couldn't build one for this work. This would have been ideal both in a qualitative approach to research and in an exploratory study as I defined it before.

Additionally, the misunderstandings in the word "Topic Maps" – which is frequently associated to "concept maps" or other kinds of visual representation of information – generated noise during the research. In the same manner, the user-display in many Web sites that claim to use Topic Maps often don't differ from others that use a different model or technology, which presented difficulties in the identification of actual applications of the standard.

Finally, one of the most representative sources of error in this research is the lack of possibilities to compare Topic Maps with models that provide similar solutions to the problems that Topic Maps addresses, such as RDF. To face this limitation, I made the basic assumption that Topic Maps has some advantages and ways to do things that are not possible to do otherwise, and that Topic Maps and its family of standards provide all the mechanisms to accomplish its main propositions.

Lastly, the ethical considerations of presenting a 'state of the art' on the placement of a concept, model or technology into a broad and changing discipline such as LIS is something implicitly unachieved. This purpose made the author of this thesis hesitate in stating opinions and undertake a frenzied search that obviously couldn't be finished.

5.2 Additional remarks on Topic Maps in LIS

5.2.1 Global identity

Identification in Topic Maps (section 4.1.1.1.) can be done on a global scale. In this case, the *subject identifiers*, when unambiguously referenced for the purpose of interoperability, are

known as *published subjects*. When they are shared for use through the Web, their respective subject indicators are known as "*published subject indicators*" or PSIs (Pepper, 2008b).

MLA are by nature institutions meant to exist in cooperation and thus, attempts to identify bibliographic entities on a global scale have gone on for years. For instance, just observing the suggested values for the element 'Identifier' of the MODS standard (hdl, doi, isbn, isrc, ismn, issn, issue number, istc, lccn, local, matrix number, music publisher, music plate, sici, uri, upc, videorecording identifier, stock number[31]), the Authority Record Number (ARN) used for authors, or the Library of Congress Control Number or LCCN, are a few examples of the long history of identification mechanisms attempted within the LIS domain.

There doesn't seem to be a way to achieve uniformity, and this is perhaps not desirable. Even though the mechanisms for common understanding are a need for cooperation and sharing, keeping the differences and attending to the needs of particular user groups is the primary mandate of each MLA institution in the current context. Both are part of the same mandate.

However, at a first glance, the mechanisms proposed for *merging* in Topic Maps seem to claim uniformity and single views. For instance, the subject identifiers, as defined in the TMDM, are the basis for merging. And the basic constructs of Topic Maps claim for 'unambiguous constructs' as it was seen in the introductory concepts.

The idea, however, behind the use of *PSIs* is a possibility for the negotiation of meaning: even though the base for *merging* is the *subject identifier*, these don't have to be unique, and people can decide which *subject identifier* would be more suitable (according to different authorities, i.e. a subject identifier from The Library of Congress or a subject identifier by another national system). Since the *subject indicator* is shared, it is possible to decide which *subject identifier* suits the needs of an institution, or of the *merging* during certain conditions or specific purposes.

The identification problems in LIS represented in the 'document-centric view' could give some clues to understand how to apply the principles of subject indicators: "the assignment of standard numbers, particularly by publishers, is often too casual to permit their being used to organize information effectively" (Svenonius, p.98). In a Topic Maps view, the standard numbers assigned by publishers are only one among several different possibilities to choose a

[31] http://www.loc.gov/standards/mods/mods-outline.html#identifier

PSI, and the interested parties in interoperating would only have to decide on the basis of which identifier the merging will take effect. On the other side, the publisher of the *PSI*, in this case the 'publishers' would have to describe, to 'disclose' in the words of Steve Newcomb (Newcomb & Biezunski, 2003), in their scope, why it unambiguously identifies the subject in question. Unambiguity is also a matter of scope.

The Topic Maps community and family of standards provides the mechanisms for this negotiation of meanings and (selection) of identities. Until now some best practices can be observed in the "Core subject identifiers" presented by the Standard for its identified terms, Ahmed (2003) with its "design patterns" which also provide the *PSIs* that would allow the sharing and reuse of representational structures (and ontologies): The Published Subject Indicators recommendation from OASIS, the Ontopedia PSIs service, and the recently launched Subj3ct: A Subject Identity Resolution Service.

Merging is the reason why Topic Maps was created, initially on a small scale, but soon realized as a mechanism that could scale. Since 2000 when it became a standard, the principles of global identification were an inherent part of its idea, and usually it is inseparable (although is possible to do it in practice) from associating it with applications at a global scale. Sigel (2000) expressed it better:

> "Topic Maps could be helpful, because they allow to define structures independent from and across the original documents, they support a more formal definition, they are open for alternative views, and they make collaborative work on evolving structures possible.
>
> Thus Topic Maps could be one apt IT that fits into Krause's layered model of information provision in which no longer a central agency exerts its authority in subject indexing and vocabulary control upon agencies located lower in the hierarchy, but in which a group of partners co-operate. Such a strategy does not result in uniform metadata, but leads to layers of heterogeneous metadata with different quality control procedures. Intellectually controlled high-quality subject schemata lie in the heart of those layers." (Sigel, 2000, p.8)

Topic Maps is a model that allows collaboration compared to other models which allows transfer.

5.2.2 Subject-centric computing

Garshol (2004) and Pepper (2008b) envision a user sitting down in front of an interface, able to start queries to it by the question 'what objects are about subject X?' (Garshol, 2004, p.?) rather than by opening specific applications or documents, or searching for specific titles and/or authors. In computing, this view has been called "subject-centric computing". There are for example subject-centric wikis (topincs), and subject-centric blogging, as well as proposals to create subject-centric personal information management systems or desktops (Pepper, 2008d).

This view has also been described as a *paradigm shift* in computing ("subject-centric computing"), similar to that represented by object-oriented programming in the 1960s and '70s (Pepper, 2008d), and as a "new Web paradigm" (Luckeneder, Steiner and Wöβ, 2001).

According to Pepper (2008d) the document-centric approach is represented in the Memex idea of Vannevar Bush, who despite his insight that information needs to be organized "as we may think" (i.e. associatively, in terms of concepts), proved unable to envisage a truly concept-centric technology. Bush's ideas inspired the field of hypertext and its visionaries (including Doug Engelbart, Ted Nelson, Bill Atkinson, and Tim Berners-Lee, the inventor of the World-Wide Web); but the Memex itself, according to Pepper, led them away from the subject-centric alternative.

In the vision of subject-centric computing, the structures used to organize documents, such as folders, hierarchies, and description, are changed for structures and applications that allow the searches to begin with concepts and facts which have related information and objects (among them documents) to inform the results. Besides this application of the idea of Topic Maps to computing, there are also different ways to apply it to Social software and Web 2.0 initiatives. Fuzzzy.com is the most cited example.

5.2.3 Semantic Web

The literature on Topic Maps and LIS seems to agree on the fact that Topic Maps is one of the enabler technologies to achieve the idea of the "Semantic Web". For instance Adams (2002), Bokman (2006), and many others, give Topic Maps such a role. Fith (2002) for example states that "One of the core ideas behind the Semantic Web is the creation of machine-

processable relationships between resource identifiers (URI's). Two often discussed ways of representing those relationships are RDF and Topic Maps."

However, this identification is problematic. On the one side, the Semantic Web –SW– with capital letters is one of the activities of the W3C, an international consortium that develops standards and guidelines for the Web. On the other hand, Topic Maps is an ISO standard whose purpose is to represent information about the structure of information resources (ISO13250). Both ideas began in different communities, at different times and for different purposes.

However, both RDF (the Semantic Web syntax) and XTM (the Topic Maps syntax) became official as respectively a W3C recommendation and an ISO standard in the same year, and their similarities didn't pass unnoticed by the two communities. Pepper (2008c) tells the story of this link and how at some point choosing one over the other was a desired purpose for the future of the Web.

However, Topic Maps and the Semantic Web (this relation is referred often as 'Topic Maps and RDF') have a different scope and are made for different purposes, among one of them, to serve for organizing information in the Web. RDF as well can be used for non Web related purposes, and that's the reason why both Topic Maps and RDF are usually encompassed with the term "Semantic web technologies".

What is behind this encompassing term is an idea about the need to apply to the Web some Information Organization principles that would allow more structured searches and results, as opposed to the searches done through word-based and ranking algorithms. Tim Berners-Lee, the creator of the W3C and the person considered to be the inventor of the Web, published in 2001 an article in the journal *Scientific American* where he said that "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

Here is where Topic Maps fit in: the idea of the Web that can use ontology-based systems to give structure to the Web. Since structuring information is one of the main purposes of Information and Knowledge Organization, the LIS community has started to adopt those 'semantic Web technologies' for these purposes, and also the RDF and Topic Maps community have looked at its historical principles for applying them in their visions.

Both the approaches to solving this vision and the vision of a semantic Web itself are problematic and have been criticized: Shirky, 2003 and Veltman, 2004, for instance. Research on the semantic Web abounds in the different communities as well as within the LIS community, for example, on its role and the implication of these technologies for its theories and practices.

Even though the differences between RDF and Topic Maps were not within the scope of this thesis, it is important to remark that they are highly interoperable due to task forces that have worked for this purpose.

What the literature on Topic Maps in LIS seems to agree upon (and this is explained by the fact that the literature comes either from the Topic Maps community or by people in the LIS community who have adopted Topic Maps) is that Topic Maps has advantages over RDF (Garshol (2002; Yi, 2008; Tramullas & Garrido, 2006; Oh, 2009 and Tuhoi, 2005). The main reason for this consideration is basically the approach to *subjects* instead of documents, and the inherently richer semantics of Topic Maps as a model over RDF due to its simplification of relations and identity mechanisms that are considered problematic in representing knowledge. Direct vs. indirect addressing seems, though, to be the main differential characteristic.

Tramullas & Garrido (2006) for instance, decided to adopt Topic Maps because of its "structure and syntax [are] more modern" and because it is a more "flexible and abstract paradigm"(p.2). These authors found in the development of their application (Potnia) that even though Topic Maps and RDF are interoperable in the sense that it is possible to represent RDF structures through Topic Maps, the other way (representing Topic Maps into RDF) represents a loss of the semantics. The authors don't specify the details of this conversion, but the literature on the interoperability of Topic Maps and RDF gives account of the reasons.

Tuhoi (2005) considered also that "Topic Maps are "higher-level" than RDF, including a few extra features". Oh (2009) concluded in his prototype development and study that "TM implementation is relatively easy compared with RDF/OWL so one can expect a better return on investment."

To conclude, one remarkable conclusion that has implications for Information and Knowledge Organization was made by Yi (2008) while comparing Topic Maps and RDF/OWL in relation to their possible application domains:

> "While RDF/OWL is optimal for making inferences about information, Topic Maps is better for finding information. RDF/OWL is suitable for the physical sciences or biomedical domains, where terms are less ambiguous; however, as terms from the humanities for social sciences have multiple meanings, making inferences using RDF/OWL is not an easy task. Topic Maps is therefore more appropriate for nonscientific fields than is RDF/OWL because Topic Maps can represent multiple meanings for each term and can build complex relationships among terms. (Yi, 2008)

However, one of the limitations of the application of Topic Maps in MLA, instead of RDF, is that precisely because the focus of Topic Maps is not specifically the Web, the solutions and implementations are often reduced to specific domains or not widely distributed and disseminated through the network. That explains the reasons why initiatives such as Linked Data appeal more to the LIS community. However, the Topic Maps community is starting to develop similar strategies to make its solutions and perspectives widely accessible. For instance, at the time of the writing of this thesis, The Ontopia Knowledge Suite (OKS) for creating topic maps was under discussion to become open source.

The **Simple Knowledge Organization System (SKOS)** is an application of the Semantic Web technology RDF to knowledge organization systems (KOS). However, in the literature on Topic Maps in LIS there no existing research was found on the comparison of Topic Maps and SKOS from an Information and Knowledge Organization perspective.

Some authors refer to the inherent advantages of SKOS. For instance, Sigel (2006) says that "SKOS provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies" (Sigel, 2006). Sigel defines all these "concept schemes" as semiformal ontologies that can be represented through the use of SKOS, in which purpose would be "to bring the worlds of library classification and Web technology together" (p.?). However, these statements are also valid if applied to Topic Maps and more research on their similarities and differences is required.

SKOS is being used already in libraries, for instance at the Deutsche Nationalbibliothek, where it was used to represent the Dewey Classification System.

There is a Topic Maps ontology for monolingual thesauri based on SKOS (Pepper, 2007).

6. Conclusions

6.1 Summary and main findings

The purpose of this research was to present a 'state of the art' of the placement of Topic Maps in Library and Information Science through an extensive literature review, creating at the same time a synthesis of their main concepts and approaches from a Knowledge and Information perspective, represented by Elaine Svenonius' *Theoretical Foundations of Informaiton Organization* and some of the concepts of Knowledge Orgnization. This thesis intended also to present both a conceptual and a theoretical framework for future research.

The qualitative study was undertaken with a Grounded Theory approach to concept analysis. One of the main elements of the research process was the creation of a conceptual framework, which served as the central point of reference for mapping terminologies and building interpretations.

The literature reviewed consisted of more than sixty documents, which included, among others, journal articles, conference presentations and papers, student reports and thesis, and a book chapter. Besides, this was complemented with information obtained from mailing lists, blog postings and websites.

As observed in this literature, Topic Maps happens to be the trigger of many issues that are the current concern of the LIS community. However, not enough research was found exploring how the standard has been incorporated into the concepts, models and practices of the LIS discipline. It showed, however, that the Knowledge Organization discipline is the main one in considering Topic Maps from a theoretical perspective, integrating it into the Information Science tradition. On the other hand, the Topic Maps literature claims that the Standard presents capabilities to be positioned at a higher level of abstraction and application than the traditional systems in Knowledge and Information Science.

The main principle upon which Topic Maps claims to differ from the bibliographic tradition is their 'identity-based' mechanisms (URI) as opposed to the 'name-based' mechanisms used in LIS for achieving identity and collocation. However, the principles behind these claims were not found to differ radically in their purposes. The basic difference is that their respective mechanisms belong to different "ages" of technology and respond to different needs given by

the context of the digital environments. Besides, since Topic Maps is at the level of abstract models, its openness regarding the specific vocabularies make possible the reuse and integration of representations that are at a lower level of abstraction, and that are based on "authority name" mappings.

This abstraction and its full adaptation to digital environments and the Web allows it to be considered as a model that builds on principles of 'integration', 'cooperation' and 'sharing' both on small and on a global scale throughout the Web. The URI identification mechanism of Topic Maps is what makes possible the representation, extension, but mainly the integration of separate bibliographic systems such as thesauri, classification schemes, metadata schemes, and term lists. This capability facilitates the integration of documents, work languages, document languages and conceptual frameworks into a single mechanism where every element represented can become an access point, a point of connection with other elements and an "information resource" on its own.

This encyclopedia-like characteristic of Topic Maps, as a superimposed information layer over documents and their representations in the LIS realm, would make it appear as a very promising technology. However, there was no strong evidence that the LIS international bodies that dispose rules and standards are considering Topic Maps in their concerns or programs of bibliographic control.

On the other hand, various applications show that Topic Maps is suitable for the representation of KOS, for integration of ILS through vocabularies, for FRBRization of library catalogs and, especially, in the field of digital libraries in the Digital Humanities for representing TEI-encoded texts.

The most representative experiences in those areas seem to be the prototype developed at the National Library of Korea for FRBRizing its MARC-based catalog, The creation of a digital library for the Finnish National Gallery, the metadata integration achieved at the folklore collection of the Department of Greek Literature at the University of Athens in Greece and the prototype for the integration of separate library systems through their vocabularies at the Danish National Library of Arts and Architecture are also promising developments.

However, the New Zealand Electronic Text Centre (NZETC) together with The Swinburn Project at Indiana University is the most consolidated application of Topic Maps found in this research. Both are in the field of Digital Humanities and have used Topic Maps together with TEI-encoded texts, for many different purposes. The projects are consolidated and mature, and seem to have guaranteed their sustainability and growing by showing positive results. Topic Maps seems to present significant advantages for these specific and semantically rich domains, since the elements can be carefully constructed and intellectually mapped.

Subject guides are reported as a promising area of small-scale applications, and the integration of MLA as the most ambitious one. The latter one is considered possible through a topic map that addresses the metadata interoperability problem at different levels. The main obstacles for these applications to be implemented seem to be in the limitations of existing records, and in the implications of intellectual requirements of identification based on meaning, as in the MLA integration project.

From an LIS perspective, Topic Maps appears to be a development aligned within the tradition of Knowledge and Information Organization but completely adapted to the context of the Web and the digital environments, not as one of the traditional Knowledge Organization Systems or its evolution, but as a bibliographic meta-language, able to not only to represent them, but to extend them and integrate them all.

This places Topic Maps on the borders of the LIS discipline with Knowledge Representation and Computer Science, where LIS conceptual models play the role of intermediaries by providing the ontologies to represent all the ‘knowledge representations’ of the bibliographic universe. As with any other technology, the human and intellectual effort is (still) needed. But the associativeness of the Topic Maps model seems to be closer to an associative way of thinking that documentalists and librarians (in different perspectives though) have had through the years.

Elaine Svenonius said that one of the central aims of her book was “to integrate the disparate disciplines of descriptive cataloguing, subject cataloguing, indexing and classification” (Svenonius, 2000, p. xi). She presented the intellectual foundations for such integration, but didn’t consider how this could be done in the context of the Web and digital information. Topic Maps, as a model, shows related interests. Besides, as a family of standards that are based on XML (the bibliographic syntax of the Web), implementation of this aim is possible.

6.2 Further research

Exploring the placement of Topic Maps in LIS has shown that it is not just an application or a tool for solving clearly defined problems or requirements and that, on the contrary, it triggers a discussion of those problems and challenges inherent in the way they are approached and reflected.

For that reason, the work and research on Topic Maps in LIS is just starting and most authors point in directions that are research areas on their own (Yi, 2008, for example, presented an inventory of these areas). This fact seems to imply that there is a need for a theoretical framework to guide and also integrate the disparate research and working efforts on Topic Maps in LIS.

Prof. Joseph Tennis recently presented to the KO interested community a classification of Knowledge Organization research situated in a meta-theoretical framework Tennis (2008). This framework, according to the level Topic Maps occupies in LIS (as it can be seen in the concept map) would be suitable within which to place the many concerns that arose with the potential of being future areas of research both at a theoretical and practical level.

As Prof. Tennis noticed, perhaps the need for a conceptual organization like this one doesn't appear to be so obvious, but I think that it will make possible an overview of the issues of concern on the placement of Topic Maps in LIS and to identify research needs. It is of course, as Prof. Tennis also stated "a naïve classification (Beghtol 2003), one created in order to demonstrate extant knowledge, with the hope creating new knowledge as a byproduct." (Tennis, 2008, p.103).

Prof. Tennis also concluded with the fact that KO discourses overlap and diverge with similar ones, such as ontology engineering and Information Architecture. His proposal for a KO research framework is divided among epistemology, theory, and methodology plus three spheres of research: design, study, and critique. Following there is a chart on the research topics on Topic Maps and LIS in the framework designed by this author:

Table 2: Theoretical framework for research on Topic Maps in LIS

Tennis (2008)	Theoretical approach	Specific approach
01 Epistemology 011 Epistemic Stance 012 Knowledge Claims 013 Assertions about Reality 014 Our Ability to Know Reality 015 What Knowledge of Reality Means 016 Acceptable Sources of Evidence in Creating Knowledge 017 Acceptable Findings	- The Epistemological assumptions of TMs - Theoretical roots of TM - TM and representation of knowledge - TM and language	- Knowledge and knowledge creation in TM - The TM approach to language - Semantics and construction of Meaning in TM - Topic Maps and Knowledge Management
02 Theory	- Topic Maps as a "paradigm shift" (shift theories)	- "Document centric" vs. "subject centric" - Subject centric computing
03 Methodology 031 Techniques 032 Contingencies to Epistemology	- TM and KOP	- Ontology design with TM - Classification with TM - Indexing with TM - Cataloging with TM
04 Design 041 Contingencies to Epistemology	- Representation of bibliographic languages with TM (design patterns) - How to build KOS with Topic Maps - How to build bibliographic systems with TM - How to create a tm in MLA - How to enhance bibliographic systems with TM - Integrating bibliographic languages KOS with TMs - Integrating bibliographic systems with TM	- (Representation, building and integration of special KOS and bibliographic systems)

	- Small scale applications of TM to MLA - Enhancement of specific features of bibliographic systems with TM	
05 Study 051 Analytical Study 052 Empirical Study	- Study of TM in relation to similar models - TM and LIS conceptual models - TM for the Digital Humanities - TM and Personal Information Management	- Evaluation of existing topic maps - TM and institutional repositories - TM applications for VLE - TM and bibliographic systems data models - Authority control in a TM perspective - TM usability studies
06 Critique 061 Critical Theory and Post-Structuralist Critique 0611 Identity in KO 0612 Work and Labor in KO 062 Discourse Analysis at Interstices of KO and Cognate Research	- TM terminology - TM and International Librarianship - TM and (universal) bibliographic control - TMs in the discourse of the Semantic Web - TM in the discourse of "social tagging"	- Comparative study of TM terminology with LIS terminology - TM and identification in LIS - TM and RDF

Some examples of research topics of TM in LIS[32] based on the previous chart:

01 Epistemology

- TM relation with the schools of epistemic thought in KO
- TM and the Referential Theory

[32] Note: 03, 04 and 05 can be applied to specific cases (institutions, systems, to especial 'bibliographic languages', to different domains, etc.)

02 Theory

- The concept of « document » in a TM perspective
- The concept of "subject" and "aboutness" in a TMs perspective
- TM and KR paradigms (e.g. conceptual graphs)

03 Methodology

- Methods for ontology design with TM
- Automatic and manual classification in TM
- Implications of a TM perspective for cataloging, indexing and classifying

04 Design

- Modeling special KOS with TM (e.g., Agrovoc Thesaurus, UNESCO thesaurus)
- Migration of existing metadata from information systems to TM
- Faceted classification with TM
- TM ontology creation based on especial bibliographic languages (DDC, thesauri, etc.)
- Use of *scope* for multilingual applications
- Uses of merging for metadata and semantic interoperability (federated searches with TM)
- TM use for subject guides and bibliographies
- Browsing and visualization of search results with TM
- "Glossing features" in topic maps[33]

05 Study

- Comparative usability studies
- TM and SKOS
- TM and TEI
- FRBRization of MARC (DC) based bibliographic systems with TM
- Comparative studies of Topic Maps-based ontology information retrieval systems with RDF-based ontology information retrieval systems 34 for bibliographic systems

[33] Dalmau & Walsh, 2007

- Comparative study of specific projects on authority records sharing and how they could be solved using RDF or TMs. Ex. NACO, VAF
- Relationships and roles in KO and KR in a TM perspective
- Hierarchical structures in KO and TM
- Integration of MLA in a TM perspective
- TM in the family of LIS standards
- Querying with TM

06 Critique

- TM, global identification and *merging*
- WorldCat in a TM perspective
- New methods of cooperative cataloging in a TM perspective
- RDA in a TM perspective

[34] Myongho Yi is the main researcher in this area.

7. References[35]

Aalberg T., Haugen, F.B., & Husby, O. (2006a). A tool for converting from MARC to FRBR. In Proceedings from ECDL'2006: *Research and Advanced Technology for Digital Libraries*: *10th European Conference* (pp.453-456). Berlin; New York: Springer.

Aalberg, T., Husby, O, Haugen, F.B., & Ore, Ch. (2006b). FRBR i bibliotekkataloger. (BIBSYS Bibliografisk gruppe). Norway. Retrieved from http://www.bibsys.no/files/pdf/publikasjoner/prosjektrapporter/prosj.rapp_frbr_i_bibliotekkataloger.pdf

Aalberg, T. (2005). *Formats and FRBR Catalogues – Where's our focus?* [PowerPoint slides]. Presented at FRBR in 21st Century Catalogues: An Invitational Workshop, OCLC. United States. Retrieved from http://www.oclc.org/research/events/frbr-workshop/presentations/aalberg/FRBR_workshop_Aalberg.ppt

Aalberg, T. (2003). *Supporting Relationships in Digital Libraries*. Doctoral thesis, Norwegian University of Science and Technology, Trondheim, Norway. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-116.

* Adams, K. (2002). The Semantic Web: differentiating between taxonomies and ontologies. *Online, 26* (4), 20-23.

* Ahmed, K. (2003). *Topic Map Design Patterns For Information Architecture*. Retrieved from: http://www.techquila.com/tmsinia.html.

* Ahmed, K. (2007). *IndiePix: A Novel approach to Topic Map Visualisation for Cultural Artifacts*. Presented at the International Topic Maps Users Conference in Oslo. Retrieved from http://www.topicmaps.com/tm2007/ahmed.pdf.

Ahronheim, J.R. (2002). *High-Level Subject Access Tools and Techniques in Internet Cataloging*. Binghamton, N.Y.: Haworth Information Press.

[35] All web addresses were last accessed on 2009-06-15. The references marked with an astherix correspond to some of the literature reviewed in Chapter 3, and make a bibliography on Topic Maps and LIS.

Aitchison, J., Gilchrist, A., & Bawden, D. (2000). Thesaurus construction and use: a practical manual. London : Aslib.

Amann, C., Beeri, I., Fundulaki, I., & Scholl, M. (2002). *Ontology-Based Integration of XML Web Resources*. Paper presented at the ISWC2002: *International Semantic Web Conference*. Retrieved from http://cedric.cnam.fr/PUBLIS/RC327.pdf

ANSI/NISO Z39.19-2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.

Auillans, P., Ossona de Mendez, P., Rosenstiehl, P. and Vaant, B. (2002). A formal model for Topic Maps. In Proceedings from ISWC: *First International Semantic Web Conference* (69-83). Berlin: Springer.

* Baratne Hajdu, A. (2005). Focus on the information. *Konyvtari Figyelo, 51* (4), 847-856.

Batley, S. (2005). Classification in Theory and Practice. Oxford: Chandos Publishing.

Berg, E. (2004). *Implementing FRBR: A comparison of two relational models: IFLA's FRBR model and Taniguchi's expression-prioritized model*. Master's thesis, Høgskolen i Oslo, Norway. Retrieved from http://www.frbr.org/2007/09.

Blumauer, A., & Pellegrini, T. (2006). Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen". In T. Pellegrini & A. Blumauer (Eds.), *Semantic Web: Wege zur vernetzten Wissensgesellschaft* (pp. 9-25): Springer.

Bock, B. (2008, January 24). Topic Maps as a Paradigm Shift. Message posted to topicmapmail mailing list, archived at http://www.infoloom.com/pipermail/topicmapmail/2008q1/007106.html

* Bøckman, K. A. (2006). *Topic Maps som browsemodel for humanistiske bibliotekers informationsformidling: Med særligt henblik på humanisters emnevokabular: En case-undersøgelse af Danmarks Kunstbibliotek*. Bachelor's thesis, Danmarks Biblioteksskole.

* Bøckman, Knut Anton. (2007). Browsing i bibliotekskataloger ved hjælp af Topic Maps. *Dansk Biblioteksforskning 3* (2), 33-44.

Bowers, S. and Delcambre, L. (2000). Representing and Transforming Model-Based Information. In Proceedings from the *International Workshop on the Semantic Web* held in conjunction with ECDL.

Bradley, J. (2003). Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools. *Literary & Linguistic Computing, 18* (2), 185-207.

Broughton, V. (2007?). *Building a faceted classification: Fundamentals of facet analysis* [PowerPoint slides]. School of Library, Archive & Information Studies, UCL.

* Bruket, B. (2008). *Automatisk FRBRisering og generering av emnekart: med et blikk på musikkmetadata.* Bachelor's thesis, Oslo University College, Oslo, Norway.

Campbell, D.G. (2006). A phenomenological framework for the relationship between the Semantic web and user-centered tagging systems. In Furner, J. & Tennis, J.T. (Eds.), *Advances in classification research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop. Austin, TX.*

Candela, L.; Castelli, D.; Ferro, N.; Ioannidis, Y.; Koutrika, G.; Meghini, C., et al. (2008). *The DELOS Digital Library Reference Model - Foundations for Digital Libraries. Version 0.98.* Retrieved December 25th, from http://www.delos.info/index.php?option=com_content&task=view&id=345#reference_model

Carlyle, A. (2006). Understanding FRBR as a Conceptual Model FRBR and the Bibliographic Universe.*Library resources & technical services, 50* (4), 264-273.

* Chen, Y., & Chen, S. (2001). An analytical review of current methodologies of knowledge organization: a knowledge discovery approach. *Bulletin of Library and Information Science, 39*, 36-51.

Chan, L., & Zeng, M. (2006a). Metadata Interoperability and Standardization - A Study of Methodology, Part I: Achieving Interoperability at the Schema Level. *D-LIB Magazine, (12)* 6. Retrieved December 21, 2008, from http://www.dlib.org/dlib/june06/chan/06chan.html

* Colmenero Ruiz, M. J. (2005). Introducción al modelo Topic Maps (ISO /IEC13250:2003). *Revista Digital de Biblioteconomia e Ciência da Informação, 3* (1), 77-102.

* Cronin, B. (2007). Annual Review of Information Science and Technology (Vol. 41). Medford, NJ: Information Today.

Dahlberg, I. (2007). Feature. Interview with Ingetraut Dahlberg. *Knowledge Organization, 35* (2/3), 82-85.

*Dalmau, M. and Walsh, J.A. (2007). *White Collaborative Award Application 2007-2008*. Bloomington, IN, United States: Indiana University, School of Library and Information Science.

* Dalmau, Michelle and John A. Walsh (2005). *Topic Maps in Digital Humanities: Sure They're Cool, but Do They Make a Difference?, A User Study*. Paper presented at the Digital Resources for the Humanities (DRH) conference, 4-7 September 2005. Retrieved from.

Davis, R., Shrobe, H. & Szolovits, P. (1993). What is a Knowledge Representation? *AI Magazine, (14)* 1:17-33. Retrieved from: http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html

Dijk, P. V. (2003). *Introduction to XFML*. Retrieved from: http://www.xml.com/pub/a/2003/01/22/xfml.html

Doerr, M. & Le Boef, P. (Eds.). (2007). *FRBR object-oriented definition and mapping to FRBR-ER (version 0.8.1)*. Retrieved from www.ifla.org.sg/VII/s13/wgfrbr/FRBR_oo_V.0.8.1c.pdf

Elings, M. W. & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. *First Monday, 12* (3). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1628

* Farquhar, A. and Bandholtz, T. (2003). Semantic Network Services Sharing an integrated Ontology using Topic Maps and Web Services. Paper presented at the *6th NKOS Workshop at ACM-IEEE Joint Conference on Digital Libraries (JCDL)*. Retrieved May 4, 2009, from http://nkos.slis.kent.edu/DL03workshop.htm

Feather, J. and Sturges, P. (2003). Information Science. In *International Encyclopedia of Information and Library Science*.

Fischer, D. (1998). From thesauri towards ontologies? In W.M. El.Hadi (Ed.), *Structures and relations in Knowledge Organization* (18-30). Würzburg: Ergon-Verlag.

* Fitch, K. (n.d.). *Taking RDF and Topic Maps seriously – what happens when you drink the Kool Aid*. Retrieved April 12, 2009, from http://ausweb.scu.edu.au/aw02/papers/refereed/fitch2/paper.html.

* Folch, H., Lahlou, S., & Benoît, H. (2000). Navigable Topic Maps for overlaying multiple acquired semantic classifications. *Markup Languages: Theory & Practice, 2* (3), 269-280.

* Garcia-Quismondo, M. A. M., & Gonzalez, J. A. M. (2001). Modelos teoricos y elementos funcionales para el analisis de contenido documental: definicion y tendencias. [Theoretical models and functional elements for analysing document content: definitions and trends]. *Investigacion Bibliotecologica: Archivonomia, 15* (31), 125-162. (In Spanish)

Garshol, L.M. (2007). Introduction to Topic Maps [PowerPoint]. Presentation given at Bouvet.

* Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all. *Journal of Information Science, 30* (4), 378-391.

Godby, C. J., Miller, E. J. & Reighart, R. R.(2001). Automatically Generated Topic Maps of World Wide Web Resources. *Journal of Library Administration 34* (3), p. 395-407.

Göktürk, E., Rosvall, I., and Gutierrez, N. (2004). *Multi-Perspective Multi-Hierarchical Topic Maps; Ontology and a Method. (Research Report, 311)*. Oslo, Norway: University of Oslo, Department of Informatics.

Guescini, R. (2006). Getting back on the trail, combating information overload with Topic Maps. Bachelor's thesis, University Of Oslo, Faculty of Humanities, Department of Linguistics and Scandinavian Studies, Norway.

Hart, C. (2001). Doing a Literature Search: A Comprehensive Guide for the Social Sciences. Sage Publications Ltd.

Hart, C. (2004). Doing Your Masters Dissertation. Sage Publications Ltd.

Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization. International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation, 35(*2/3), 86-101.

Hjørland, B. (2006a). *Lifeboat for Knowledge Organization*. Retrieved from http://www.db.dk/bh/lifeboat_ko/home.htm

* Hjørland, B. (2006b). Topic Map. *Lifeboat for Knowledge Organization.* Retrieved April 12, 2009 from http://www.db.dk/bh/lifeboat_ko/CONCEPTS/topic_maps.htm

Hjørland, B. (1992). The Concept of subject in Information Science. *Journal of Documentation, 48*(2), 172-200.

* Howard, K. J. (2008). An expression of FRBR in Topic Maps. Unpublished paper.

IFLA (1998). IFLA Study Group on the functional requirements for bibliographic records. *Functional requirements for bibliographic records* (final report). Munich, Germany: K. G. Saur. Retrieved from http://www.ifla.org/VII/s13/frbr/frbr.pdf

* Iglesias, E. & Stringer-Hye, S. (2008). Topic maps and the ILS: an undelivered promise. *Library Hi Tech, 26*, p. 12-18. (Emerald)

ISO 21127:2006. *Information and documentation: A reference ontology for the interchange of cultural heritage information.*Retrieved from http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.1_Mar09.pdf

ISO/IEC 13250-1: 2003. *ISO/IEC 13250 Information Technology—Topic Maps—Part 1: Overview and Basic Concepts*.

ISO/IEC 13250-2: 2006. *ISO/IEC 13250 Information Technology—Topic Maps—Part 2: Data Model*.

ISO/IEC CD 18048. *ISO/IEC CD 18048 Information Technology—Topic Maps—Constraint Language (TMCL)*. Retrieved from http:// www.isotopicmaps.org/tmcl/.

* ISO/IEC DTR 29111: 2007. *ISO/IEC DTR 29111 Information technology – Topic Maps – Expressing Dublin Core Metadata using Topic Maps*. Retrieved from http://www.jtc1sc34.org/repository/0884.htm.

Johannesen, A. (2004, April 5). *Thoughts on Topic Maps*. Posted to http://shelter.nu/ blog/

* Johannesen, A. (2006, May 26). *The epistemological implications of Topic Maps for librarians*. Posted to http://shelter.nu/blog/2006/05/epistemological-implications-of-topic.html.

* Johannesen, A. (2006, September 11). *Topic Maps.* Message posted to Next Generation Catalogs for Libraries. Retrieved from http://listserv.nd.edu/cgi-bin/wa?LIST=NGC4LIB

* Johannesen, A. (2007, November 23). Explaining Topic Maps to Library Scientists [Msg 2]. Message posted to http://ligent.net/pipermail/topicmapsinlis/, archived at http://ligent.net/pipermail/topicmapsinlis/2007-November/000013.html

* Johannesen, A. (2008). *(Fever) visions for a Topic Mapped library*. Presentation given at the International Topic Maps Users Conference in Oslo. Retrieved from http://www.topicmaps.com/tm2008/johannesen.pdf

* Johannesen, A. and Pearce, J. (2004). *Topic Maps Awareness Seminar* [PowerPoint slides]. National Library of Australia. Retrieved from *shelter.nu/blog-096.html*

* Kivela, A., & Lyytinen, O. (2004). *Topic Map aided publishing: A case study of Assembly Media Archive*. Retrieved March 18, 2009, from http://www.wandora.net/wandora/wiki/images/Topic_map_aided_publishing.pdf.

* Kivelä, A., & Lyytinen, O. (2007). *Case study: publishing large collection of artworks using Topic Maps*. Presentation given at the International Topic Maps Users Conference in Oslo. Retrieved from http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2009&id=Aki_Kivela~TM2007~Publishing_Large_Collection_of_Artworks.

* Kongsbakk, I. (2003). *Emnekart og tesauri – gjenfinning i nettverk og hierarkier.* Thesis for Diplomoppgave ved Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag.

* Kongsbakk, I. (2004). Sømløs kunnskap: om bruk av emnekart. *ABM-skrift, 8.*

* Laursen, H., & Henriksen, M. (2006). *Integrating Multiple Library Catalogues.* Presentation given at the International Topic Maps Users Conference in Oslo.Retrieved from http://www.topicmaps.com/tmc/presentation.jsp?conf=TM2006&id=Henrik_Laursen~TM2006~Integrating_Library_Catalogues.

* Le, D. & Nguyen, Q. (2007). *Search Framework for a large digital records archives*. Presented at the Digital Library Federation Spring Forum. United States. Retrieved from http://www.diglib.org/forums/spring2007/presentations/nguyen.pdf.

* Lee, H., Jeon, Y., & Sung-Kook, H. (2006a). MARCXTM: Topic Maps Modeling of MARC Bibliographic Information. In *Maicher & J. Park (Eds.), Charting the Topic Maps research and applications landscape (Vol. 3873, pp. 241–252). Berlin/Heidelberg, Germany: Springer*.

* Lee, H., Jeon, Y., & Sung-Kook, H. (2006b). *MARCXTM: Topic Maps Modeling of MARC Bibliographic Information*. [PowerPoint slides] Presented Charting the Topic Maps Research and Applications Landscape, TMRA 2005, October 6-7.

Leise, F.; Fast, K. and Steckel, M. (2003). *Controlled Vocabularies: A Glosso-Thesaurus*. Retrieved from: http://www.boxesandarrows.com/view/controlled_vocabularies_a_glosso_thesaurus

* Leuenberger, M., Grossmann, S., Stettler, N., & Herget, J. (2006). Topic Maps for image collections. In L. *Maicher & J. Park (Eds.), Charting the Topic Maps research and applications landscape (Vol. 3873, pp. 107–111). Berlin/Heidelberg, Germany: Springer*.

* Liukkonen, E. (2008). *The Collections web service of the Finnish National Gallery*. [PowerPoint slides]. Presented at Arlis Norden 2008 conference. Retrieved from http://www.arlisnorden.org/island/radstefna2008/eija_pdf.pdf

* Lixin, X., Jin, Z., & Zhongyi, W. (2008). Cross-language information retrieval model of digital library based on topic map. *Journal of the China Society for Scientific and Technical Information, 27*(3), 468-473. Retrieved from www.csa.com

* Lourdi, I., Christos Papatheodorou & Ch., Nikolaidou, M. (2007). A multi-layer metadata schema for digital folklore collections. *Journal of Information Science, 33* (2) 2007, p. 197–213. Retrieved from EBSCO.

Lor, P.J. (2008). *An introduction to International Librarianship* [PowerPoint slides]. Presented at his lecture at the University of Florence and University of Parma.

Luckeneder, T., Steiner, K, and Wöβ, W. (2001). *Integration of Topic Maps and Databases: Towards Efficient Knowledge Representation and Directory Services*. DEXA, LNCS 2113, p.744-753.

* Marzal G., M. A., Beltran O., P., Morato L., J. L., Moreiro G., J. A., & Sanchez C., S. (2002). Propuesta para la utilización de estructuras verbales aplicadas a la recuperación y representación de la información. [Proposal for applying verbal structures in information representation and retrieval]. *Scire, (8),* 95-102.

* Meschini, F. (2005, May). Le mappe topiche: come imparai a non preoccuparmi e ad amare i metadati [Topic maps How I learned to stop worrying and love metadata]. *Bollettino AIB: Associazione Italiana Biblioteche*, (1), p. 59-72.

Morville, P. (2007). Information Architecture for the World Wide Web. O'Reilly.

* Newcomb, S. (2006). *Evolutions in Subject Searching: the use of Topic Maps in Libraries.* Presentation at the American Library Assocation LITA Forum, October 27, 2006, Nashville, Tennessees.

Newcomb, S. R. (2003). A Perspective on the Quest for Global Knowledge Interchange. In: *J. Park & S. Hunting (eds.). (2003). XML Topic Maps: Creating and Using Topic Maps for the Web. Boston: Addiso-Wesley*.

* Newcomb, S.R. and Biezunski, M. (2003). *Topic Maps, the inventors perspective on subject access*. Presented at the Library of Congress Luminary Lectures series. Retrieved from: http://www.loc.gov/rr/program/lectures/mbsn.html.

NISO Z39.19-2005. *NISO Z39.19-2005: Guidelines for the construction, Format, and Management of Monolingual Controlled Vocabularies*.

* Norrish, J. & Stevenson, A. (2008). *Topic Maps and Entity Authority Records: an Effective Cyber Infrastructure for Digital Humanities*. Presented at Digital Humanities, 2008. University of Oulu, Finland.

* Oh, S. G. (2008a). *Evaluating an Innovative Korean Folk Music (Pansori) Retrieval System Using Topic Maps*. Retrieved March 31, 2009, from http://www.garshol.priv.no/tmcase1/presentation.jsp?id=presentation-TopicMaps2007-14.

* Oh, S. G. (2008b). *MARC, FRBR and RDA: The Topic Maps Perspective.* Presentation given at the International Topic Maps Users Conference in Oslo.

* Oh, S. G. (2008c). *Topic Maps in Library Science (tutorial: Topic Maps for Library and Information Scientists)*. Presentation given at the International Topic Maps Users Conference in Oslo. Retrieved from http://www.topicmaps.com/tm2008/oh2.pdf

* Oh, S. G. (2007). *Topic Maps-Driven Semantic Services for National Library of Korea. Kioto, Japan.* Presentation given at the International Topic Maps Users Conference in Oslo. Retrieved from http://www.knowledge-synergy.com/atoms/slides/atoms2007-samoh2.pdf.

Okada, T. et al. (2007). Classification and Organization of Medical words using Topic Maps. [PowerPoint slides]. Presented at AToMS 2007. Retrieved from www.knowledge-synergy.com/atoms/slides/atoms2007-okada.pdf.

Park, J., & Hunting, S. (Eds.) (2002). *XML Topic Maps: Creating and Using Topic Maps for the Web*. Boston: Addison-Wesley Professional.

Passin, T. B. (2004). *Explorer's Guide to the Semantic Web* (illustrated edition.). Manning Publications.

* Peng, Y., & Ke, H. (2008). The Study of Using Topic Map in Pathfinder. *Journal of Library and Information Science*, *34*(1), 39-61.

Pepper, S. (2000, rev. 2002). *The TAO of Topic Maps: Finding the way in the age of information*. Retrieved from http://www.ontopia.net/topicmaps/materials/tao.html.

* Pepper, S. (2007). *MyThesaurus: A Topic Maps ontology for monolingual thesauri based on SKOS*. Retrieved from http://onlywright.ligent.net/omnigator/models/topicmap_nontopoly.jsp?tm=MyThesaurus.xtm&pageView=themes

* Pepper, S. (2008a). Expressing Dublin Core in Topic Maps. In *Scaling Topic Maps, Lecture Notes in Computer Science* (pp. 186-197). Retrieved January 25, 2009, from http://www.ontopedia.net/pepper/papers/DCinTopicMaps.pdf.

* Pepper, S. (2008b). Topic Maps. *Encyclopedia of Library and Information Sciences*. [forthcoming]

Pepper, S. (2008c). *Topic Maps and The Semantic Web*. Posted at http://topicmaps.wordpress.com/

Pepper, S. (2008d). *Everything is a subject*. Closing keynote at the International Topic Maps Users Conference in Oslo. Retrieved June 6 2009 from http://www.topicmaps.com/tm2008/pepper.pdf.

* Pepper, S., Sigel, A., Schmidt, I., & Müller, C. (2000). *Topic Maps for Knowledge Organization and Knowledge Management.* Presentation at ISI: 7th International Symposium for Information Science. Germany.

* Pharo, N. (2004). Topic maps – knowledge organisation seen from the perspective of computer scientists. Paper presented at European Library Automation Group (ELAG) conference. Norway. Retrieved from http://www.elag2004.no/papers/Pharo.pdf

* Pharo, N. (2008). Emnekart og kunnskapsorganisering – ny flaske for gammel vin? [PowerPoint slides]. Presented at Nolug (Norsk online brukergruppe). Oslo, Norway. Retrieved from http://biblioteknett.no/nolug/emnekart.pdf.

* Ramalho, J., Librelotto, G., & Henriques, P. (2006). Metamorphosis—A Topic Maps based environment to handle heterogeneous information resources. I In L. *Maicher & J. Park (Eds.), Charting the Topic Maps research and applications landscape (Vol. 3873, pp. 14-25). Berlin/Heidelberg, Germany: Springer*.

Seaman, C.B. (2007). Empirical paradigm: Position paper. In *Empirical software engineering issues: Critical assessment and future directions. International Workshop, 26-30 June 2006* (pp.1-3). Berlin, Heilderberg: Springer.

Schmitz-Esser (1999): Thesaurus and Beyond: An Advanced Formula for Linguistics Engineering and Information Retrieval. *Knowledge Organization, 26,*10-22.

Shirky, Clay (2003, November 7). *The Semantic Web, Syllogism, and Worldview.* First published on the "Networks, Economics, and Culture" mailing list. Retrieved from http://www.shirky.com/writings/semantic_syllogism.html

* Sigel, A. (2006, Octubre 10). *Knowledge Organization with Topic Maps-Revisited*. Leipzig, Germany. Retrieved from http://asigel.de/uploads/media/TMRA06_Tutorial_Knowledge_Organization_v16_2006-10-10_01.pdf.

* Sigel, A. (2004). *FRBR and XTM* . Retrieved from http://web.archive.org/web/20060425125810/http://kpeer.wim.uni-koeln.de/~sigel/Projects/FRBR_and_XTM.html

* Sigel, A. (2003). Topic Maps in Knowledge Organization. In *Park, J., & Hunting, S. (Eds.) XML Topic Maps: Creating and Using Topic Maps for the Web (pp. 383-476). Boston: Addison-Wesley Professional.*

* Sigel, A. (2000). *Towards knowledge organization with Topic Maps* [Power Point slides]. Presented at XML Europe 2000. France. Retrieved January 22, 2009 from http://www.gca.org/papers/xmleurope2000/papers/s22-02.html

Sowa, J.F. (2000a). *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole: Pacific Grove, CA.

Sowa, J.F. (2000b). Ontology, Metadata, and Semiotics.In *B. Ganter & G. W. Mineau, (eds.), Conceptual Structures: Logical, Linguistic, and Computational Issues, Lecture Notes in AI #1867, Springer-Verlag, Berlin, 2000, pp. 55-81.* Retrieved from http://www.jfsowa.com/ontology/ontometa.htm

Sowa, J.F. (1992). Semantic Networks. *Encyclopedia of Artificial Intelligence* (S. C. Shapiro, Ed.), NY: Wiley, New York, 1987; revised and extended for the second edition, 1992. Retrieved from http://www.jfsowa.com/pubs/semnet.htm

* Stevenson, A., & Styron, E. (2006). *Topic Map Presentation Framework: an Approach to Delivering Newspaper Content Over the Web*. Retrieved February 18, 2009, from http://researcharchive.vuw.ac.nz/handle/10063/63?show=full.

* Stevenson, A., Darwin, J., & Tuohy, C. (2006). *Going Beyond Google: Representation and Retrieval of Information Using Topic Maps*. Retrieved February 18, 2009, from http://researcharchive.vuw.ac.nz/handle/10063/137.

* Stevenson, A., Tuohy, C. & Norrish, J. (2008). *Ambient findability and structured serendipity: enhanced resource discovery for full text collections*. New Zealand Electronic Text Centre. Retrieved from http://hdl.handle.net/10063/267

* Stringer-Hye, S. (2005). Reinventing the wheel: can librarians help? *OCLC Systems & Services: International digital library perspectives (21)* 2. p. 89-91.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. United States: The MIT Press; 1 edition.

Svenonius, E. (2004). The Epistemological Foundations of Knowledge Representation. *Library Trends, (52)* 3, 571-587. Retrieved from http://findarticles.com/p/articles/mi_m1387/is_3_52/ai_n6080404/?tag=content;col1

Tennant, R. (2002). MARC must die. *Library Journal*. 27, 26-27. http://www.libraryjournal.com/article/CA250046.html

Tillet, B. (2003?). *What is FRBR*. Retrieved from http://www.loc.gov/cds/downloads/FRBR.PDF

* Tramullas, J. & Garrido, (2006). Constructing Web subject gateways using Dublin Core, the Resource Description Framework and Topic Maps. *Information Research (11)*2. Retrieved from http://informationr.net/ir/11-2/paper248.html

Tudhope (2006). *A tentative typology of KOS: towards a KOS of KOS?* Presented at ECDL06: 5th European NKOS Workshop, European Conference on Digital Libraries, Alicante, Spain.

* Tuohi, C. (2005, May 5). *Topic Maps @ NZETC*. Retrieved January 25, 2009 from http://www.nzetc.org/tm/scholarly/tei-NZETC-About-technology.html

* Tuohi, C. (2008). *Topic Maps for Cultural Heritage Collections* [Power Point slides]. Presented at International Topic Maps Users Conference in Oslo. Retrieved from http://www.topicmaps.com/tm2008/tuohy.pdf

* Tuohy, C. (2007). *Topic Maps and TEI: Using Topic Maps as a Tool for Presenting TEI Documents*. Retrieved February 18, 2009, from http://researcharchive.vuw.ac.nz/handle/10063/160?show=full.

* Vassallo, Salvatore (2006). Navigating through archives, libraries and museums: Topic Maps as harmonizing instrument. In In *Maicher & J. Park (Eds.), Charting the Topic Maps research and applications landscape (Vol. 3873, pp. 231–240). Berlin/Heidelberg, Germany: Springer*.

Veltman, Kim H. (2004). Towards a Semantic Web for Culture. *Journal of Digital Information, (4)4.* Retrieved from http://jodi.tamu.edu/Articles/v04/i04/Veltman/veltman.pdf

* Walsh, J. (2005). *TM4DH (Topic Maps for Digital Humanities): Examples and Open Source Toolkit*. Presented at ACH/ALLC. Retrieved February 17, 2009, from http://mustard.tapor.uvic.ca/cocoon/ach_abstracts/xq/pdf.xq?id=169.

* Walsh, John A. and Michelle Dalmau. (2006). *Navigating a Sea of Texts: Topic Maps and the Poetry of Algernon Charles Swinbure*." Presented at the Digital Library Federation Spring Forum. The Driskill Hotel, Austin, Texas. 10-12 April 2006.

* Weston, P. G. (2002). Dal controllo bibliografico alle reti documentarie. From bibliographic control to documentation networks. *Biblioteche Oggi.*

* Yi, M. (2008). Information organization and retrieval using a topic maps-based ontology. *Journal Of The American Society For Information Science And Technology, 59*(12), 1898-1911.

Yi,M. (2006, September). Ontology as a technology of convergence: Developing a classification of the ITWF using Topic Maps. Paper presented at the annual conference of the Association of Internet Researchers (AoIR), Brisbane, Australia.

Appendices

I. List of initial codes

The following list presents alphabetically some of the basic topics (or codes) that emerged from the literature in the step 5 of the data analysis (section 1.3.5). Some of them are taken as they were in the texts, but mostly, while reading, I named them with a different term trying to map the concepts according to my initial knowledge. They were subsequently grouped into “code families” that became (through interaction while building the conceptual framework) the categories and concepts that are presented in Chapter 2.

- AACR2
- Access point
- Cataloguing
- Classification schema
- Controlled vocabulary
- Design pattern
- Descriptive cataloging
- Dewey Decimal Classification
- Digital library
- Document
- Document centric
- Dublin Core
- Entity-relationship
- Faceted classification
- Findability
- FRBR
- FRBRization
- Hierarchies
- Identifiers
- Indexing
- Information Architecture
- Information Retrieval (IR)
- Information resource
- ISBD
- Knowledge organization (KO)
- Knowledge representation
- LCSH
- Library catalog
- Library technique
- MARC
- MARCXML
- MARCXTM
- Metadata interoperability
- MODS
- Ontology
- Property-value pairs
- Resource
- RDA
- RDF
- Semantic interoperability
- Semantic Web
- SKOS
- Social tagging
- Standard
- Subject
- Subject cataloging

- Subject centric
- Subject heading
- Synonym ring
- Taxonomy
- TEI
- Thesaurus
- Topic Maps
- UDC
- XML
- XSLT
- XTM

II. Summary of applications of Topic Maps to MLA and the Digital Humanities

The following is a summary of the applications found in the literature. These corresponds to a**pplications that claim to use topic maps**.

Catalogs of libraries, archives, museums

- Danmarks Kunstbiblioteks katalog (prototype not implemented) (Denmark)
- Fundación Germán Sánchez Ruipérez. Uses "Tema Tres", supported by Topic Maps, in their digital library. (Spain)
- Korean National Library (prototype in implementation) (Korea)
- Stanford University Libraries' HighWire Press: http://highwire.stanford.edu/help/hbt/ (browsing aid based on indexing terms). (USA)
- The Royal Library (prototype not implemented) (Denmark)

Digital collections in the Humanities

- Arppeanet (Finland)
- Arthurian Legend. http://www.letrs.indiana.edu/tm4dh/topicmaps/arthurianLegend.xtm (USA)
- Assembly Media Gallery 2004-2005 (Finland)
- Auslit. http://www.austlit.edu.au/
- (Note: it is inspired by Topic Maps but not run by a topic map). (Australia)
- Cedeca. http://www.cedeca.it/progetto.aspx
- (Note: about 1600 topics and 4000 associations) (Italy. University of Pavia)
- Classical Mythology. http://www.letrs.indiana.edu/tm4dh/topicmaps/mythology.xtm (USA)
- Collections of Finnish National Gallery http://www.fng.fi (Finland)
- Fish trout, you're out (prototype) (Australia)
- Kiasmart (Finland)

- Korean Folk Music (Pansori) Retrieval System (Korea)
- Korean Historical resources (Korea)
- Mauritius Collection (prototype not implemented). (Australia)
- Mediateekki & Media Archive. (Finland)
- Musica Migrans. http://www.musicamigrans.de/. (Germany)
- New Zealand Electronic Text Centre (NZETC) at Victoria University of Wellington http://www.nzetc.org/ (New Zealand)
- The Emigration Museum. www.museu-emigrantes.org (Portugal)
- The folklore collection of the University of Athens Greek Literature Department (Greece)
- The Living Memory (in progress) (Germany)
- The Swinburne Project (USA)
  http://www.letrs.indiana.edu/tm4dh/topicmaps/swinburne.xtm
  http://swinburnearchive.indiana.edu/swinburne/www/swinburne/index.html
- Town again (Note: taken offline due to copyright issues) (Finland)
- WebKat.hu http://www.webkat.hu/scripts/webkat (Hungary)

Newspaper content delivery in libraries

- Topic Maps Presentation Framework for newspaper content delivering (New Zealand, Australia)
- OmniPaper (Canada) http://canada.esat.kuleuven.ac.be/omnipaper/

Other uses of Topic Maps in MLA

- Hálózatos Irodalom (Hungary)
  http://mekmester.oszk.hu:8080/itm/tmv/index.htm
- (E-learning application on Hungarian literature for secondary school students created by the National Library of Hungary)
- Picture Australia: National treasures http://nationaltreasures.nla.gov.au/ (Australia)
  (temporary Website for a traveling exhibition)
- Potnia (tool for creating subject pathways using Topic Maps, DC, RDF) (Spain)
- Chung Hua University Library (pathfinder). (China)

- Tema Tres (a tool for creating thesaurus –out of the scope of this thesis though, it is for Information Architecture) http://tematres.r020.com.ar/index.html (Argentina)

Feasibility studies reported in papers or mailing lists

(information on conclusions is not available, and it's not official)

- National Archives and Records Administration (NARA). www.diglib.org/forums/spring2007/presentations/nguyen.pdf (USA)
- National Library of New Zealand
- National Library of Norway