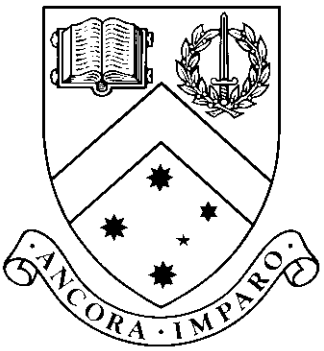
School of Computer Science and Software Engineering
Monash University



Literature Review — Semester 2, 2004

ConceptBib: A Tool for Integrating Concept Maps and Bibliography Management

Phan Anh Le, 1802 8136

Supervisor: Dr. David Squire

Contents

1 Introduction

There is an increasing need for an effective way of organizing knowledge in order to optimize its use. Organizing knowledge involves storing information in suitable medium, structuring information in various ways, and appropriately delivering information to users so that efficient knowledge retrieval, proliferation are possible. An example of situations, where knowledge management (KM) is essential, is the literature review process. It is necessary that doing literature review requires a lot of reading, note taking, and thinking in order to weigh the encountered materials. Moreover, what is vital is to identify the key ideas and the relationships between them, which can be massively complex in relatively large research projects. The challenge is to manage research information properly from which new knowledge can be derived as well as correct references must be made.

There have been various effort put forward in researching of sound KM, amongst of those is using concept maps (CM) to structure information into the graphical form. More specifically, graphical form or graphs, show advantages in representing the relationships amongst illustrated entities; in particular case of KM, they can show the semantic structure of the knowledge. CMs can allow us to simplify complex knowledge for which the use of serial textual documents is considered inadequate. This literature review briefly introduces concept maps and its applications in bibliographic management.

Looking from a more practical point of view, there are applications that can speed up, as well as increase the power and usefulness of KM process. In the particular case of doing research, they are tools that assist researchers in writing papers in general, including typesetters (e.g. LaTeX) and bibliographic management helpers such as BibTeXor EndNote. Despite the advantages these tools provide, the tasks involving doing research review, such as effectively managing ideas and their references, remains a challenge and needs further improvement, which may require a combination of various techniques from different scientific domains. This literature review aims at describing and evaluating a number of existing techniques and tools used in general knowledge management and scientific research, and it also proposes a combined method that can bridge the gaps between them.

2 Review of Concept Map

This section briefly describes the theoretical foundation of CMs, which originated in psychology fields. This sections also looks at the applications where CMs have perform well, and introduces some of their practical tools.

2.1 An overview of concept maps

Jonassen et al. (1993) defined concept maps as "representations of concepts and their interrelationships that are intended to represent the knowledge structure that humans store in their minds". In plain language, concept maps are tools for better management of human knowledge, which help users learn new information more quickly and effectively as they appear in graphical form.

In a less abstract description, a typical concept map is a graph, which consists of two sets: concepts, often represented by named nodes, and the relationships between them or propositions shown by connecting labelled arcs. In essence, the concepts are

the generalizations of knowledge, of ideas conveyed in any form, e.g. books, documents, speeches, lectures, or observations whereas propositions show how concepts are linked together (Novak; 2002). Intuitively, through a quick look at the map, the authors or the new users would have much more convenience in studying and remembering the semantic contents of the map. The psychological explanation is that human brain learns and stores information much more efficiently in the graphical forms.

The figure 1 is an example of CMs, which is a visual description of how a simple computer can be constructed. It is clear that this pictorial representation can be used in replacement of a textual paragraph detailing the required computer components. Without loosing the essential details, the semantics of the computer composition is equivalently captured within a simple yet information-rich concept map.

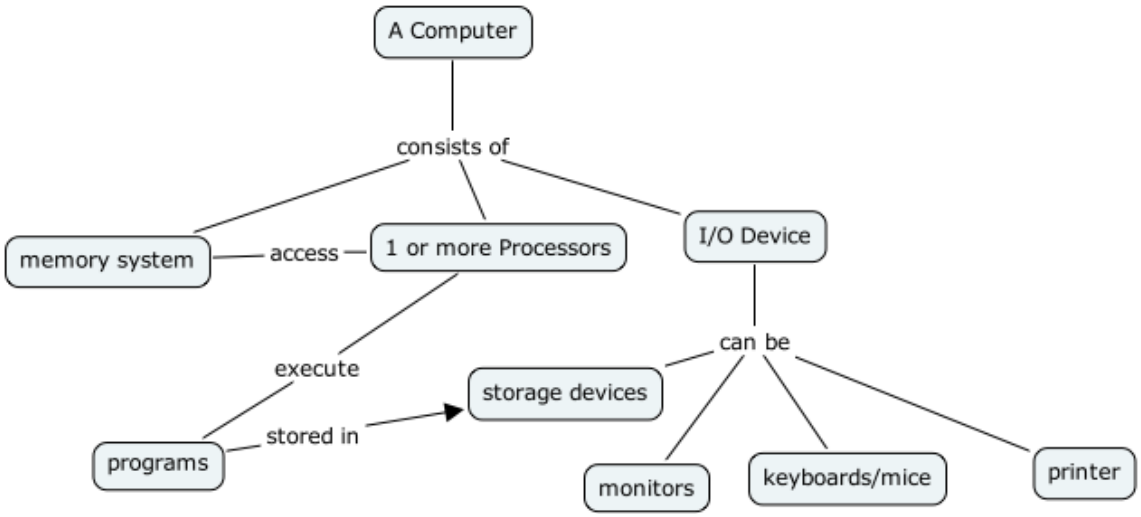


Figure 1: Simple components of a personal computer

To talk briefly about the CMs history, they were introduced as a part of a larger project back in 1960s where researchers analyzed the development of children's knowledge. In that research, Ausuel et al. (1978) proposed that learning is the integration of new knowledge into learners existing framework, which serves as the foundation for thinking and action. Children start to form their cognition from early ages. As they grow up, their frameworks are gradually extended, deepened, and corrected since their preliminary perceptions are not always true. In addition, in order to be stored long term memory, new knowledge must be connected to others via relationships rather than being an isolated entity, which can be easily forgotten (Ausuel et al.; 1978).

To model the process of building up a knowledge framework, and to follow the Ausubels research, the definition of concept maps was proposed (Novak; 1998). In this theory, building up a knowledge framework is considered to be similar to the creation of a concept map from scratch. The knowledge has inter-relationships, and so do the concepts. By drawing upon good concept maps, educators can benefit in teaching and evaluating what learners have gained from their education (Novak; 1991).

2.2 Applications of Concept Maps:

Although concept maps originated in psychology, they have been used widely in business, engineering, and particularly education. Here we present examples of areas in which concept maps have shown their strength:

- Facilitating Collaborative Learning: In many areas, efforts invested in learning and creating new knowledge do affect the outcomes. Concept maps can be used in efficiently learning new knowledge. If the shared concept maps are created by groups of learners, they have also been useful in knowledge exchange across individuals as well as larger groups (Novak; 2002).

- Assistance in Knowledge Creation: Because of intuitive nature of concept maps, businesses and engineering corporations have used concept maps in brainstorming new ideas, roadmaps of productions, relationship diagrams, many of which have not been formally recognised as concept maps (J.P. van Schie; 2002).

- Concept Maps for Evaluation: Not only do the concept maps show their effectiveness in knowledge proliferation, they are also a great tool for evaluation of retained information by the learners (Edwards and Fraser; 1983).

2.3 Review of current tools applying Concept Maps

A number of concept map-based applications exist such as MindMapping (The Mindtools Webteam; 2004), IHMC Concept Map Tools (The IHCM Group; 2004). In regard to the context of this project, we are especially interested in MindMapping and IHMC Concept Map Tools because of their popularity and their provided functions .

2.3.1 MindMapping technique and its related software

MindMapping is a commercial trademark of Buzan Organisation (Buzan Webteam; 2004). It is a systematic technique complemented with provided commercial software used mainly for purpose of quick note-taking and summarization. MindMapping advances the way that people take notes by providing a 2-dimensional work space, where users can easily create "shapes" of subjects and important links between them. Mind maps are generally compact than conventional list notes, and it retains essential information in an intuitive way.

MindMapping graphs can be quickly created and extensible they have predefined templates and shapes. MindMapping shows its usefulness in integrating information from various sources, summarizing information, and representing the overall structure of the domain (The Mindtools Webteam; 2004).

2.3.2 IHMC Concept Map Tools

The Institute for Human Machine Cognition (IHMC) Concept Map tool is the product of the research organization where the definition of concept maps themselves was coined. The IHML tool is a general purpose tool with no predefined templates, however it empowers users to easily create, navigate, share and criticize concept maps amongst joint research community.

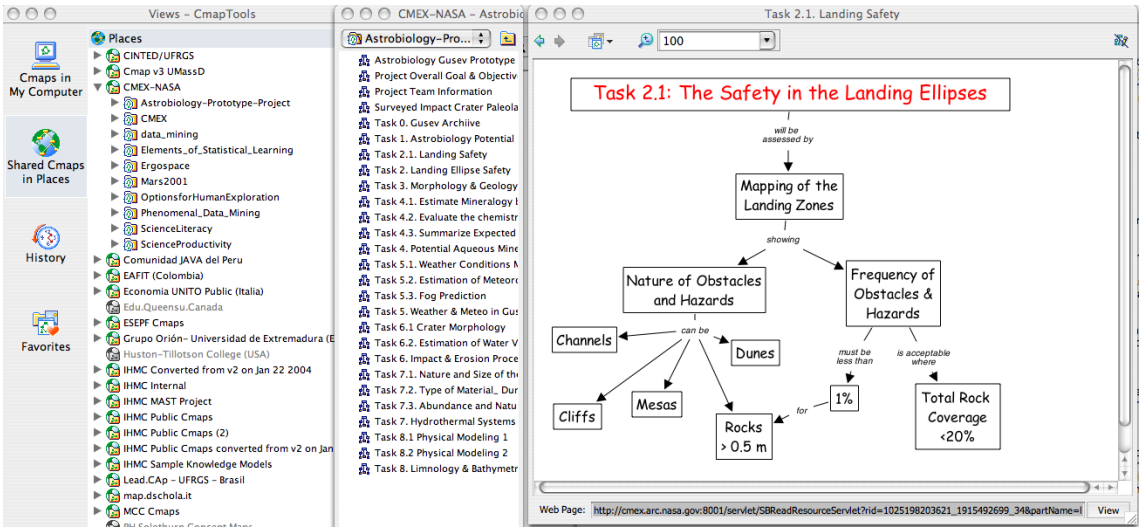


Figure 2: Connect the IHCM tool to pre-registered NASA sites to retrieve an existing Concept Map

The strongest advantage of this tool is that it can connect to various registered research sites and retrieve existing concept maps stored in those remote locations. The figure 2 illustrates this function in graphical user interface. This is beneficial if the authors want to search, refer, or integrate the current map into others already shared and standardized by the wider community. In addition, IHMC tool is platform-independent as it is written in Java. The users can create, modify, and export the concept maps to different formats including image files (such as JPEG, TIFF, and PNG), web pages, and XML file formats at ease.

However, there are still limitations that can be improved from the existing software:

- Limitation in referenced resources: the strong point introduced by IHMC tool is also its limitation. It allows users to connect only to trusted pre-registered hosts, and reference to files stored in those hosts or file store locally. In other words, it limits the number of resources the current concept map can point to.

- Limitations in information embedded in nodes: As it is a general concept maps, users find little support in nesting further information into the nodes. In the current version, it only allows users to enter a few lines of text which will be displayed when users scroll the mouse over specific nodes.

- Limitation in grouping and layering: Current version provides no support for layering or grouping certain types or groups of concepts. These function can be useful in a more complex concept maps where searching and layering are important.

In summarizing the review of current tools, there are a number of current limitations in both commercial and educational software that this project aims at partly addressing:

- Hyperlinks and resource locators not fully supported: In the Internet age, it is desirable to be able to quickly access the resources and information supporting a

specific concept, not only stored locally or in limited predefined sites. A better solution might include the degree of relevance to the current concept so that the users can select the best suited when they needed. Interaction with other software such as document readers or web browsers would also be useful options.

- Embedded information is limited: Embedded information to concept might be more than simple text. It might include key words used by search engines, references to other resources, multi media files such as images, or audio files. Generally speaking, the embedded information can be pointers to anything that can be used to find supporting or contrasting information of current concepts.

- Classification or layering are not fully supported: As the knowledge grows, graphical representation can be increasingly complex. A function allowing users to quickly identify and group parts of the concept maps would be required.

3 Representation of Bibliographic and Content Metadata

This section is dedicated to looking at management of bibliographic information that can be embedded into the concept maps described in previously. Bibliographic information is generally used to help users to find the exact resource. This section describes and discusses various methods used to represent bibliographic data.

3.1 BibTEX and its current limitations

3.1.1 BibTeX—A brief introduction:

The creation of TEX by Donald Knuth in the 60s marked the new era of the typesetting of documents, particularly mathematical and technical documents (Kopka and Daly; 1999). TEX offers primitive commands that deal with the simplest formatting functions. Also provided along with TEX is a set of high quality fonts, and even its own processing language (Knuth; 1986).

To continue the success of TEX, the successor, LATEX, which was developed by Leslie Lamport in 1985, is more user-friendly than TEX. While TEX technically focuses on the formatting and typesetting of documents by issuing low level commands, the main goal of LATEX is to give a set of higher level functions, which are easier for users to use and memorize (Lamport; 1986). LATEX provides markup commands allowing users to easily produce chapters, sections, figures, tables of contents along with others powerful functionalities.

In 1985, Oren Patashnik along side Lamport developed a tool called BibTEXwhich processes bibliographic information embedded in LATEX documents. This provides users with an global mechanism of processing citations in LATEX (Kopka and Daly; 1999). Since then, LATEX and BibTEX have become amongst the most frequently used tools when researchers write academic papers, particularly in computer science and engineering domains. The mechanism in which LATEX and BibTEX collaborate is simple yet powerful: LATEX has the reference keys in its, which point to segments of textual bibliographic data in BibTEX files. After several runs, LATEX automatically looks at the segments, extracts and integrates them neatly into the referring documents.

3.1.2 Current Limitations:

Although BibTEX has greatly helped writers in processing references since it was created, as requirements of users increased and the types of publication diversified, users started to find a number of limitations in BibTEX. Current limitations of BibTEX which we have observed include:

- Serial presentation and access: Data stored in BibTEXis in plain text therefore it is serially accessed. Serial access can also means the lack of convenience and slowness.
- Lack of standards for online publications: as BiTEX was developed in pre-Internet age, it does not provided formal methods of dealing with online publications.
- Structural issues/cross reference: BibTEX provides 14 listed standard types of citations (Kopka and Daly; 1999). To use any other types, such as images or multi media files, users need to modify style files which are difficult to code, time consuming.

To solve those limitations, an increasing body of research has proposed tools and data models, most of which are eXtensible Markup Language (XML)-based (The W3C Webteam; 2004). The next coming sections describe these approaches.

3.2 A brief review of EndNote

EndNote is also a bibliographic management suite (The ISI Research Soft; 2000), which is widely used in research community on Microsoft Windows platforms. Noticeable features include:

- Graphical User Interface (GUI) -based: aiming that users who may not be familiar with text-based bibliographic tools, EndNote provides a user-friendly interface with proper layout.
- Connects directly to variety of libraries around the world: Endnote provides collaborative interfaces which enable users can search in pre-registered library collections remotely through EndNote gateways.
- Integrated with Office suites such as one of Microsoft Office: Users can point to reference items easily within the opened editors without switching to EndNote. The results later are inserted following the standard appearance.
- Supports a wide range of bibliographic formats: EndNote can import various popular file formats such as Machine Readable Cataloging (MARC)
- Strong search capability: Built in EndNote suite are the capabilities of complex search functions that allow users to locate specific resources based on various keys such as the title, authors, subjects.

However, it is observed that EndNote is a commercial product, currently mainly focusing on Windows users and Microsoft platform's software. By doing so, EndNote does not provide support to wider research community, specially computer science and engineering, who enjoy the professional layouts, fonts, and mathematical formulas provided by typesetters such as LATEX.

3.3 XML Approaches to Bibliographic Data

From the theoretical point of view, the function that BibTEX and EndNote are playing is to contain a higher level of information, that can assist the readers to quickly access the original data. Equivalently, information technology has the term *metadata*, which in plain language means data about other data. Currently, most of the prominent metadata schemes are in XML or XML related formats since they are portal, easily extensible, well structured, well understood, and well supported by applications across domains (Hillman; 2003). By drawing upon technology offered by XML based documents, we can address part of the BibTeX limitations stated previously, including the lack of standards for online based publications, and the structural and cross references issues.

This section looks at a numbers of the most popular representations that have been put forward and used widely in general knowledge management, particularly librarianship.

3.3.1 Brief look at Dublin Core

Dublin Core (DC) is the short name for The Dublin Core Metadata Initiative, a project dedicated to promoting the proliferation of interoperable metadata standards and developing more intelligent networked information discovery systems (Powell; 2003). In a much more specific terms, DC aims at developing metadata standards used in across domain, and introducing frameworks that facilitate the use of metadata standards.

Technically, Dublin Core provide a set of XML terms that can be used to effectively describe summary information of the target resource. The highest level element set includes standardized 15 items which is presented verbatim in Appendix **??** and a more comprehensive set of qualified Dublin Core term can be viewed at DC's website (Dublin Core Metadata Initiative; 2003). These elements are used to semantically cluster information of the subject into sub sections for efficient information discovery.

It is worth noticing is that DC element set is not designed for use as a static set, rather than that it provides foundations for much larger and more domain-specific sets of elements defined based on particular requirements. DC introduces an effective standardized guideline, which can be extended or used in conjunction with other body of elements.

3.3.2 Metadata Object Description Schema (MODS)

MODS is introduced by the Library of Congress (LOC) in order to provide a relatively complete set of bibliographic elements that can be used directly in various applications, particularly in librarianship (The Library of Congress; 2004). Being created after Dublin Core and in recognition of limitations of DC applications in specific domains, MODS offers a richer and more complete vocabulary. A complete list of elements and attributes of MODS can be viewed in details at LOC website at *http://www.loc.gov/standards/mods/*

Functionally, MODS can provide as many operations as DC does. Moreover, MODS facilitates the resource descriptions of not only online resources but of any type, ranging from textual contents to multi media such as images, sound recording, or mixed material. Unlike DC as a standard guideline and normally not being used as a standalone model, MODS can be used separately without the further creation of specific body of elements. This is because MODS has strong background support and interrelationships with a range of standards and resource locator systems such as Machine Readable Cata-

loging (MARC) (The Library of Congress; 2003) or Digital Object Identifier (DOI) (The International DOI Foundation; 2003).

Shown in figure 3 is a example of the BibTEXentry and figure 4 illustrates the equivalent representation in MODS.It is worth noticing that much more information, such as the subject, the genre of the described resource, resource locators (e.g. URL, DOI, or ISBN), keywords in search engines, access dates, and language information can be easily embedded into MODS, which makes this standard much more open and extensible. (The Library of Congress; 2004)

```
@book{Novak1998,
   author= {Novak, J.D.},
   title = {Learning, Creating and Using Knowledge: Concept Maps
             as Facilitative tools in School and Corporation},
   year= {1998},
   publisher={Lawrence Arlbaum and Associates},
 }
```

Figure 3: An example of BibTEX entry

```
<mods ID="Novak1998">
    <titleInfo>
        <title>Learning, Creating and using Knowledge</title>
        <subTitle>Concept Maps as Facilitative tools in School and Corporation</subTitle>
    </titleInfo>
    <name type="personal">
        <namePart type="family">Novak</namePart>
        <namePart type="given">JD</namePart>
        <role>
            <roleTerm authority="marcrelator" type="text">author</roleTerm>
        </role>
    </name>
    <originInfo>
        <dateIssued>1998</dateIssued>
        <publisher>Lawrence Arlbaum and Associates</publisher>
    </originInfo>
    <typeOfResource>text</typeOfResource>
    <genre authority="marc">book</genre>
    <identifier type="citekey">Novak1998</identifier>
</mods>
```

Figure 4: Bibliographic information in MODS

Looking at these standards within the context of this project, it would be more suitable to apply MODS as it inherits advantages of DC as well as introduces improvements. However, from software engineering point of view, it is also desirable to leave the design independent of proprietary formats. Extra consideration would need to be invested to ensure this goal.

4 Linking Concept Map and Bibliographic Management

Having discussed these two reasonably mature fields, namely concept maps in psychology and bibliographical management tools and standards in document publishing, we believe that their strengths can be further exploited if the two approaches are combined in a single tool. This tool will enable users to generate concept graphs, to affix references to supporting evidence, and to interactively update the users' reference databases such as BibTEX or EndNote.

An possible approach is that we can improve and extend existing open source software so that it provides required functionality. A potential candidate is Dia, a general purpose and popular drawing tool, which are widely used in open source community (George; 2000). Dia provides strong existing libraries, which can handle various shapes and related functions. Design and source code in Dia are extensible so that new information can be embedded as we require bibliographic data is inserted in. A complete description and review of how Dia can be employed to produce a better tool will appear in the final thesis.

References

Ausuel, D., Novak, J. and Hanesian, H. (1978). *Educational Psychology, A Cognitive View*, Hold, Reschart and Winstons.

Buzan Webteam (2004). Buzan Overview, *Buzan Website* . Access date: 10–07–2004.
URL: *http://www.mind-map.com/EN/*

Dublin Core Metadata Initiative (2003). DCMI Metadata Terms. : Accessed 10–07–2004.
URL: *http://www.dublincore.org/documents/dcmi-terms/*

Edwards, J. and Fraser, K. (1983). Concept maps as reflectors of conceptual understanding, *Research in Science Education* **13**: 19–26.

George, H. (2000). Dia Tutorial, *Dia Drawing Tool Website* . Access Date:26–03–2004.
URL: *http://www.lysator.liu.se/ alla/dia/*

Hillman, D. (2003). Using Dublin Core, *Using Dublin Core Web Site* . Access date: 26–03–04.
URL: *http://dublincore.org/documents/usageguide/*

Jonassen, D., Beissener, D. and Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying and acquiring structural knowledge*, Erlbaum, Hillsdale(N.J).

J.P. van Schie (2002). Concept Mapping Visualization Tools for Knowledge Management. Accessed 10–07–2004.
URL: *http://www.observetory.com/conceptmappingvs1.htm*

Knuth, D. (1986). *The TeXbook*, Addision-Wesley.

Kopka, H. and Daly, P. (1999). *A Guide to LaTeX: Document Preparation for Beginners and Advanced Users*, 3rd edn, Pearsion Education Ltd.

Lamport, L. (1986). *LaTeX: a document preperation system*, Addision-Wesley Pub. Co.

Novak, J. (1991). Clarify with concept maps, *The Science Teacher* **58(7)**: 45–49.

Novak, J. (1998). *Learning, Creating and using Knowledge: Concept Maps as Facilitative tools in School and Corporation*, Lawrence Arlbaum and Associates.

Novak, J. (2002). The theory underlying concept maps and how to construct them, *Website of Institute for Human & Machine Cognition (IHMC)* . Accessed Date:13–04–2004.
URL: *http://cmap.coginst.uwf.edu/info/*

Powell, A. (2003). Expressing Dublin Core in HTML/SHTML meta and link element, *DCMI Official Website* . Access Date: 26–03–2004.
URL: *http://dublincore.org/documents/dcq-html/*

The IHCM Group (2004). IHCM Concept Map Tool. Accessed 10–07–2004.
URL: *http://cmap.ihmc.us*

The International DOI Foundation (2003). DOI Overview.
URL: *http://www.doi.org/*

The ISI Research Soft (2000). *EndNote: the All-in-One solution*, Berkeley, Calif: ISI ResearchSoft.

The Library of Congress (2003). MARC Standards.
URL: *http://www.loc.gov/marc*

The Library of Congress (2004). Outline of Elements and Attributes in MODS Version 3.0. : Accessed 10–07–2004.
URL: *http://www.loc.gov/standards/mods/v3/mods-3-0-outline.htm*

The Mindtools Webteam (2004). MindMaps - A Powerful Approach to Note Taking, *MindTools Website* . Access Date: 10–07–2004.
URL: *http://www.mindtools.com/pages/article/newIS_01.htm*

The W3C Webteam (2004). Extensible Markup Language (XML). Access 26–07–2004.
URL: *http://www.w3.org/XML/*