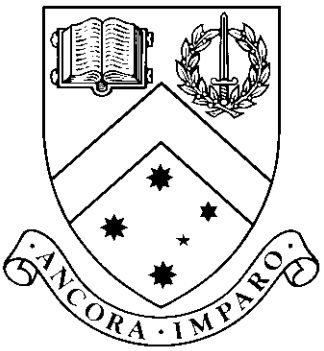
School of Computer Science and Software Engineering
Monash University



Research Proposal — Semester 1, 2004

ConceptBib: A Tool for Integrating Concept Maps and Bibliography Management

Phan Anh Le, 1802 8136

Supervisor: Dr. David Squire

Contents

1 Introduction

Doing research, particularly creating a literature review, requires researchers to consult a great deal of papers and other sources of information. They often notice key concepts or ideas in single or multiple papers. The ideas are sometimes stated explicitly via words, figures, and diagrams, or they can be also implied in higher level of abstraction, which require readers' effort to form summary, analysis, and generalization. Some questions frequently asked are [1]: [2]

- Intellectual property credits: Where did the ideas come from, are the authors the first to propose the key ideas or these has been published elsewhere, what was the domain of the ideas?
- Impact of concepts: What were other publications that lead to the ideas, and how did these ideas influence others that followed?
- Supports and challenges: What was the supporting evidence, or did any other researchers state this differently?
- Relationship to reader's work: How would these be beneficial to reader research; is this a supporting proof or a refutation?

Bearing these questions in mind, the readers then analyze and structure these key concepts into their own knowledge for later reviews and references. They commonly circle keywords, write marginal notes, or highlight the sections that contribute to the ideas. For more persistent reference, readers also keep lists of references or records of read papers, for example files in *BibTeX* or *EndNote* formats [3], as well as soft or hard copies or links to where they can found the publications for efficient access. Combining with desktop publishing tools, these bibliography management databases help them in the writing process by automatically and systematically generating references to cited papers. With the popularity of authoring and publishing formatters such as *LaTeX* or Microsoft Word, this seems to be the dominant practice in writing research papers.

There are, however, some existing restrictions in current prominent bibliography management tools. When the researchers need to comprehend and keep in mind the relations between papers and their underlying concepts, current tools offer no functionality of storing relationships between ideas which in turn are generalized from references. In that situation, they often devise their own solutions such as adding plain text summaries and notes directly to the reference database or storing them elsewhere. Nevertheless, the writing process is still difficult as they need to navigate through an ocean of textual summaries to find correct citations.

So the problem is in the textual representation of summaries and its inability to show complex interconnections among concepts and publications. In order to address this problem, psychological studies propose that visual graphs are far superior than plain text in representing complex, non-linear, cross-referenced, and relationship-based information such as the concepts in publications. Good examples include the use of formalized diagrams in software modeling, particularly Unified Modeling Language (UML) [4] where class diagrams, package diagrams and others effectively represent the relationships and interactions between software components. Exploring further into the field of psychological knowledge and learning theories, studies have proven that graphs also help readers to

learn much more efficiently and stay in long-term memory if the information is converted to graph-based representations [5].

In order to exploit the advantages of visual graphs in information representation, this project aims to look at their use for representing key concepts and relations between concepts as a solution to the problem stated above. It also has the goal of providing researchers with *ConceptBib*, an open source drawing tool, which will be able to generate concept graphs, and to interactively update the users' reference database such as *BibTeX* or *EndNote* via this tool. It is also expected that if a group of researchers has the same field of interests, concept maps of a member can be greatly beneficial to those of others in familiarizing with the context. If the project was successful, in future development a framework would be developed so that researchers could share their maps, and merge closely related ones.

2 Research Context

2.1 Concept Maps and their applications

This section briefly describe the definition of concept maps and existing tools applied this theory in various applications.

2.1.1 Concept Map

For the project to successfully address the problem, it is necessary to understand foundation theories which will be employed in this project. Fundamentally, doing a literature review is similar to a process of building up a knowledge structure of a specific domain. The researchers start with little or no basic information, then continuously construct more advanced and well structured knowledge as they progress in reading papers, introducing experiments, analyzing actual results, and writing new papers [5]. Being similar to constructing a building in civil engineering, forming a knowledge structure needs to be bases on, and enlarge, the existing foundation. Knowledge within an effective learner's mind is a complex, yet highly organized structure that allows the learner to conceptualize and access information effectively [6].

In order to model the human brain knowledge structure, the idea of representing knowledge visually as graphs is introduced in the definition of concept maps by Novak [7]. In his definition, a typical concept map is a graph which consists of two sets of components: concepts often represented by named boxes or circles, and the relationships or propositions showed by connecting labeled lines. In essence, the concepts are the generalization of knowledge, of ideas conveyed in some forms, e.g. books, documents, speeches, or lectures, whereas propositions show how a concepts are linked together. Figure 1 shows an example of simple concept map that helps a student to memorize why we have different seasons in a year.

2.1.2 Utility of Concept Map

Psychological studies demonstrate that the concept map can effectively model the development of the knowledge virtually held in human brains [6]. New information entering learners' heads need to go through a number of states: stored in short-term memory,

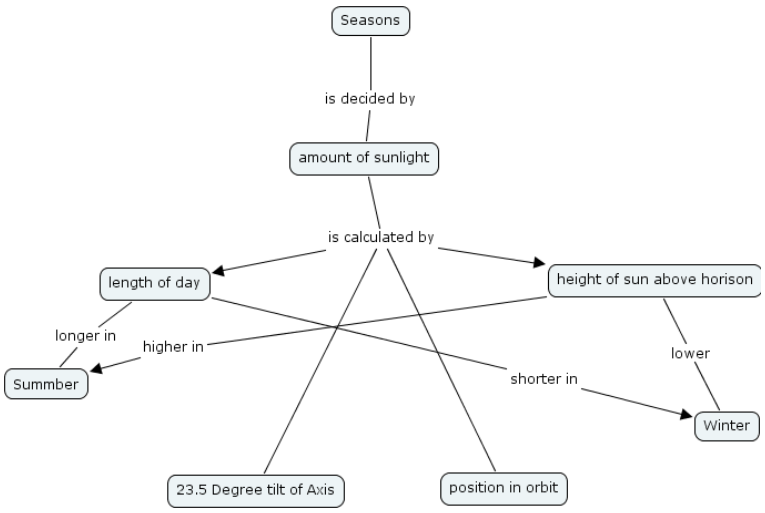


Figure 1: What decides the season of the year?

linked with the existing information structure, and then finally absorbed as new long-term knowledge. The course of meaningful learning described above is considered similar to the process of building concept maps, and consequently, we believe, similar to the literature reviewing.

2.1.3 Current Concept Maps Tools

Since concept maps were invented, they have been widely employed to help users in knowledge building process across various domains, from engineering and business to education. Consequently, there are several existing software tools to assist users to create concept maps. The most popular include *MindMap* [8], *Inspiration* [9], and the *Concept Map tool* of Novak's research center [10]. Having undergone a careful design and evaluation process, these tools show strong capacity to assist users in brainstorming, linking concepts and visual thinking. In particular, the*Concept Map tool* even allows users to attach the specific *Unified Resource Location (URL)* of supporting evidence to the maps. Although that will be greatly helpful for users to refer back the sources when needed, this does not offer full bibliographic details.

2.2 Bibliography Management Tools and Standards

This section discuss the existing bibliographic management tools and the standards in which the information is stored.

2.2.1 Existing Bibliography Management tools

Another facet of the issue that this project is considering is the variety of existing bibliographic management tools, including *BibTeX* and *EndNote*. As mentioned above, these tools are popularly used since they can accurately and systematically build up the database of references, and help writers refer back to sources when necessary. In particular, *End-*

Note, which is commercially developed, even provides the connections to various library catalogues allowing users to remotely search and quickly save the results to their own database. Furthermore, they can also interact with various type of software applications, from Microsoft Word to complex and powerful formatters such as *LaTeX* . To exploit this interoperability, the project will also look at how to produce outputs so that they are readable to these tools.

Figures 2 and 3 show two small snippets of source extracted from a *BibTeX* and a *EndNote* files which referred to a same book.

```
@book{Wordsworth1996,
author={Wordsworth,J.B},
year={1996},
title={{S}oftware {E}ngineering with {B}},
publisher={Addison-Wesley},
address={Harlow, England}
}
```

Figure 2: Typical example of *BibTeX* entry

```
<RECORD>
<YEAR>1996</YEAR>
<ISBN>0201403510 0201403560</ISBN>
<AUTHORS>
<AUTHOR>Wordsworth, J. B.</AUTHOR>
</AUTHORS>
<TITLE>Software engineering with B</TITLE>
<PLACE_PUBLISHED>Harlow, England ; Reading, Mass.
</PLACE_PUBLISHED>
<PUBLISHER>Addison-Wesley,</PUBLISHER>
<PAGES>xv, 331 p.</PAGES>
</RECORD>
```

Figure 3: *EndNote* exported *eXensible Markup Language (XML)* file

2.2.2 Other metadata standards

Given that each bibliographical management tool has its own file format (like the case of *BibTeX* and *EndNote*) and the project aims at an portable way of storing publication data, it is worth considering other existing standards to attach the desired information to the concept maps. These include the *Dublin Core Metadata Initiative* (normally referred as *Dublin Core*) [11], the increasingly popular *Digital Object Identifier* (*DOI*), and exchange format *BibTeXML*.

- **Dublin Core**: The *Dublin Core* is a well established standard for generation of metadata to describing resources [12]. A metadata set, which describe a resource

at a high level, can have different tags capturing information such as authors, title, locations, publishers, time of publications. This means that *Dublin Core* can successfully identify a resource via these descriptors. It has a strong cross-domain community support ranging from computer science to librarianship. In terms of representaion, *Dublin Core* can be represented in various formats, nowadays mainly in *XML*. The following shows an example of *Dublin Core* used to describe the same book referred in previous examples:

```
<dc:creater>
Wordsworth, J.B
</dc:creator>
<dc:date>   1996  </dc:date>
<dc:title>
    Software Engineering with B
</dc:title>
<dc:publisher>
     Addison-Wesley, Harlow, England
</dc:publisher>
```

- **Digital Object Identifier (DOI)**: Another new yet becoming increasingly popular method of identifying digital resource is using DOI [13], which is essentially a persistent unique identifier allocated to each document. The idea is similar to the ISBN in hardcopy publishing, via which a publication can be distinctively identified. Large scientific publishers such as the *Association of Computing Machinery (ACM)*, *Kluwer*, and *Elsevier Science Direct* are successfully employing this scheme to help locate publications more systematically. Once users have the DOIs of publications, they can quickly acquire others related information via online search in globally synchronized DOI databases. A quick example can be that given the `doi:10.1006/jmbi.1998.2354`, it can be instantly resolved to return the URL of the paper named *Strutural biology of HIV* [14] available online from *Science Direct*. The key point here is the persistency of information attached to the paper: the actual location where the soft copy is stored can be changed overtime, but given the permanent DOIs, it can be easily retrieved.

- **BibTeXML**: The *BibTeXML*, on the other hand, is a method of representing *BibTeX* information in *XML*. It was proposed because *BibTeX* is still prominent in the research community, and at the same time *XML* applications are becoming widespread in online publications. *BibTeXML* project also provides tools to convert data from *BibTeX* to *XML* and vice versa. Following is a quick example of a *BibTeXML* entry:

```
<BibTeX:entry id="esbensen">
<BibTeX:book>
<BibTeX:author>Kim Esbensen</BibTeX}:author>
<BibTeX:title>Multivariate Analysis in Practice</BibTeX:title>
```

```
<BibTeX:publisher>Camo AS</BibTeX:publisher>
<BibTeX:year>1994</BibTeX:year>
<BibTeX:address>Trondheim</BibTeX:address>
</BibTeX:book>
</BibTeX:entry>
```

We believe that these formats can essentially store the information considered in the context of this project. It is expected that, however, the extensions or combinations of these are possible in the course of project development, as we need to envisage a feasible, yet extendable and adaptable, method to suit existing and potential applications that this project aims to collaborate with.

2.3 Linking Concept Map and Bibliography Management

Having discussed these two reasonably mature fields, namely concept maps in psychology and bibliographical management tools in document publishing, we believe that their strengths can be further exploited if 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 database such as *BibTeX* or *EndNote*.

3 Research Plan and Methods

3.1 Methodologies

For the *ConceptBib* tool to be successful, the project aims at three stages of development: drawing tool to create the concept map, the link from concepts to references, and the interface for the concept map to interact with existing bibliographical database like *BibTeX*.

The first stage of the project is to extend or build up a drawing tool so that it can allow users to create concept maps. It will involve creating new shapes for concepts, menus, and functions for each type of shape. The new tool will allow users to create, label and connect new concepts. Additional functions include grouping related concepts, viewing concept maps in different layers. A noticeable potential candidate for the project is the open source drawing tool *Dia* [10]. *Dia* offers a variety of useful functions to create different shapes, and lines connecting them. Layers used to represents objects in different groups are similar to our aims to view concepts as different scopes. Having been in open source community for reasonably extended of time, *Dia* is well-tested and extendable for us to include new shapes, menus, and functions.

The second stage will look at the integration of bibliographical information into the graph files. This will allow users to create and affix relevant information to concepts in the map. As stated in above section, further technical analysis will be needed to weight strengths and weaknesses of each format, then select the most suitable.

The third stage will involve developing a mechanism for the drawing toolkit to interact with existing bibliographical database. This will allow direct updates in *BibTeX* or *EndNote* database when users create or edit new references in concept maps. The emphasis of this stage will be on devising ways of automatically exporting formats that *ConceptBib* produces to files formats that *BibTeX* and *EndNote* can easily import. If time permits,

this stage might consider some result analysis by giving *ConceptBib* to research students, receiving their feedbacks, and assessing the effect of the tool.

The integration of three stages will provide users with benefits in effectively learning new knowledge and managing references from which the researchers would augment the process of writing literature reviews.

3.2 Proposed Thesis Chapter Headings

1. Introduction
2. Foundation theories
3. Current standards and applications
4. Investigation on the effect of visualization
5. Integrating bibliography information to concept maps
6. Interacting with bibliographical database
7. Evaluating the ConceptBib
8. Conclusion and future work
9. Bibliography
10. Appendix A: Software developed

3.3 Timetable

This is a proposed project timeline for the rest of the semester 1 and semester 2.

Table 1: Project Timetable

Date	Task
March 17	Begin reading literature
March 24	Begin looking at *Dublin Core* and *BibTeXML*
April 1	Begin looking at *Dia* Code
April 10	Begin writing research proposal
April 15	Draft research proposal
April 28	Submit research proposal and start coding stage 1
May 4	Prepare interim presentation
May 13	Begin literature review
May 28	Start coding stage 2
June 14	Submit literature review draft
July 8	Start coding stage 3
July 29	Submit literature review
August 14	Finalise coding and start system testing and evaluation
September 2	Show draft of thesis to supervisor
September 8	Submit draft of thesis and finalise evaluation
October 1	Prepare presentation
October 17	Final presentation
November 1	Submit log book & thesis & software
November 9	Finalise project website

3.4 Special Facilities Required

All required software, hardware, and other facilities are available for use at the School of Computer Science and Software Engineering, Monash University. In the course of extending software, additional level of access might be needed to compile source code and install software in computers in the labs.

References

[1] Li Gangmin, Uren Victoria, Motta Enrico, Shum Simon Buckingham, and Domingue John. ClaiMaker: Weaving a Semantic Web of Research Papers. In *1st International Semantic Web Conference*, pages 20–30, Sardinia, 2002. ISWC2002.

[2] Ferguson Ernest. Object-oriented Concept Mapping using UML class diagrams. *The Journal of Computing in Small Colleges*, 18(4):344–354, 2003-April.

[3] ISI Research Soft. *EndNote: the All-in-One solution.* Berkeley, Calif: ISI ResearchSoft, 2000.

[4] Larman Craig. *Applying UML and patterns: an introduction to object-oriented analysis and design.* Prentice Hall, Upper Saddle River,N.J., 1997.

[5] Novak JD. *Learning, Creating and using Knowledge: Concept Maps as Facilitative tools in School and Corporation.* Lawrence Arlbaum and Associates, 1998.

[6] Ausuel D, Novak J.D, and H Hanesian. *Educational Psychology, A Cognitive View.* Hold, Reschart and Winstons, 1978.

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

[8] Concept Draw Webteam. Mindmap Overview. *Concept Map Website*, 2004. Available: http://www.conceptdraw.com/en/products/mindmap/overview.php Access Date: 25–04–2004.

[9] Inspiration Webteam. The Power of Visual Learning. *Inspiration Website*, 2004. Available: http://www.inspiration.com/vlearning/index.cfm Access Date: 25–04–2004.

[10] Harry George. Dia Tutorial. *Dia Drawing Tool Website*, 2000. Available: http://www.lysator.liu.se/ alla/dia/. Access Date:26–03–2004.

[11] Powell A. Expressing Dublin Core in HTML/SHTML meta and link element. *DCMI Official Website*, 2003. Availble at:http://dublincore.org/documents/dcq-html/ Access date: 26-03–2004.

[12] Hillman D. Using Dublin Core. *Using Dublin Core Web Site*, 2003. Available: http://dublincore.org/documents/usageguide/ Access date: 26–03–04.

[13] John Erickson. Digital Object Identifier. *McGraw-Hill Yearbook of Science & Technology 2003*, 2003.

[14] Turner Brian and Summers Michael. Structural biology of HIV. *Journal of Molecular Biology*, 285:1–32, 1999 8 Jan.