Using Concept Maps as a Tool for Cross-Language Relevance Determination

Ryan Richardson

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science and Applications

Edward A. Fox, Chair
Benjamin E. Cline
Manuel Perez-Quiñones
Naren Ramakrishnan
David Tegarden

June 6, 2007
Blacksburg, Virginia

Keywords: concept mapping, cross-language information retrieval,
computing ontologies, named entity extraction

Using Concept Maps as a Tool for Cross-Language Relevance Determination

Ryan Richardson

Abstract

Concept maps, introduced by Novak, aid learners' understanding. I hypothesize that concept maps also can function as a summary of large documents, e.g., electronic theses and dissertations (ETDs). I have built a system that automatically generates concept maps from English-language ETDs in the computing field. The system also will provide Spanish translations of these concept maps for native Spanish speakers. Using machine translation techniques, my approach leads to concept maps that could allow researchers to discover pertinent dissertations in languages they cannot read, helping them to decide if they want a potentially relevant dissertation translated.

I am using a state-of-the-art natural language processing system, called Relex, to extract noun phrases and noun-verb-noun relations from ETDs, and then produce concept maps automatically. I also have incorporated information from the table of contents of ETDs to create novel styles of concept maps. I have conducted five user studies, to evaluate user perceptions about these different map styles.

I am using several methods to translate node and link text in concept maps from English to Spanish. Nodes labeled with single words from a given technical area can be translated using wordlists, but phrases in specific technical fields can be difficult to translate. Thus I have amassed a collection of about 580 Spanish-language ETDs from Scirus and two Mexican universities and I am using this corpus to mine phrase translations that I could not find otherwise.

The usefulness of the automatically-generated and translated concept maps has been assessed in an experiment at Universidad de las Americas (UDLA) in Puebla, Mexico. This experiment demonstrated that concept maps can augment abstracts (translated using a standard machine translation package) in helping Spanish speaking users find ETDs of interest.

This research work was funded in part by NSF through grants DUE-0121679 and IIS-0080748, a grant from IMLS, and the RDEC/ACE project.

To my parents and grandparents

Acknowledgements

First I thank my Lord and Savior Jesus Christ, in whom I live and move and have my being. I also thank my adviser, Dr. Edward Fox, for never losing faith that I would actually finish this PhD, and for his endless source of new ideas about my research. I also thank my other committee members, Dr. Manuel Perez-Quiñones, Dr. Naren Ramakrishnan, Dr. David Tegarden, and Dr. Ben E. Cline for their excellent suggestions regarding research directions. Special thanks go to Dr. Douglas Oard, who was on my committee until the very end and was only removed due to travel conflicts. He offered numerous helpful critiques and made me aware of many topics in cross-language retrieval that I otherwise would have missed. His input definitely improved the quality of this dissertation.

Special thanks go to Fernando Das Neves for his translation work and proofreading for the UDLA experiment. Without his help the experiment would not have been possible.

I thank Ben Goertzel, Hugo Pinto, and Vicente Cordeiro for their help with modifying Relex. Thanks go to Lillian Cassell and her team at Villanova for their work on The Ontology Project. I thank Craig Scott for making the Scirus collection available. I also thank Alberto Castro Thompson and Silvia González Marín at UNAM for making their ETD collection available to us. I thank Alfredo Sanchez for making the UDLA ETD collection available, and (along with Gabriel Torres) for making their UDLA students available for the cross-language experiment. Thanks go to Seungwon Yang, Seonho Kim, and Doug Gorton for being our domain experts in digital libraries. Thanks to Dee Dee Dudley for handling travel arrangements while was I here--the conference travel definitely made my time here more fun. Thanks go to Uma Murthy for being willing to drive 12 hours (each way) with me on a trip to Pensacola to visit the Institute for Human and Machine Cognition (IHMC). Thanks go to Greg Hill, Larry Bunch, Tom Eskridge, Roger Carff, and all the rest at IHMC for their invaluable help interfacing with CMapTools. Finally, I thank my wife, Karen, for her love and patience while sticking with me through what seemed like the PhD which would never end. Well, the end is here, and now it's off to the next adventure!

Table of Contents

List of Figures

List of Tables

1 Introduction

"I recall attending a dissertation defense on a very similar topic. Unfortunately the dissertation is in Portuguese." -- *motivating example from Dagobert Soergel, U. of Maryland, October 2004.*

1.1 Statement of problem

The growth of the World Wide Web has led to increased availability of many large documents, such as electronic theses and dissertations (ETDs). In fact, the Networked Digital Library of Theses and Dissertations (NDLTD) [83] has made over 300,000 ETDs available in at least 12 different languages. However, it is difficult for users to determine if such large documents, many written in a language they cannot read, are relevant to their information needs—so they can seek a translation. Unfortunately, automatically translating large documents, like ETDs, so they easily can be found, read, and understood, is beyond the current state of the art in machine translation (MT), especially in technical fields.

A goal of my research is to make ETDs more readily accessible as an information resource for students and researchers. Some challenges are that ETDs tend to be very long, and often only one section will pertain to a user's information need. Also, I would like to make it easier for users to determine if an ETD, or part of an ETD, is relevant to their information need, even if they cannot read (or have difficulty reading) the language in which the ETD is written. If the user determines that the ETD is relevant, they can either attempt to read the original ETD (if they have limited language proficiency in the original language), or they can have part or all of the ETD professionally translated.

In my research, I have limited myself to ETDs about computing, allowing me to use an ontology of computing/information science that is being developed by ACM [23]. Since large ontologies are currently being developed for many scientific fields by various organizations [99], my methodology should also be applicable to other disciplines outside computer science.

My method is, by using entity-extraction combined with an ontology of computer science, to produce concept maps (a type of semantic network) of an ETD which users can browse. For cross-language use, the concept maps will be translated using MT. I contend that these automatically-generated concept maps can be used in addition to the author's own abstract to make relevance determinations easier.

1.2 Concept maps and summaries

Since the 1980's, researchers in the education field have studied concept maps as a means to facilitate the quick and effective learning of concepts [26]. Concepts are defined as "perceived regularities in events or objects, or records of events of objects, designated by a label" [86]. Concept maps, first suggested by Joseph Novak, can be defined as "graphical representations of knowledge that are comprised of concepts and the relationships between them" [87]. They consist of labeled nodes and links and are regarded as useful pedagogical tools. As envisaged by Novak, concept maps have concepts presented in an hierarchical fashion, with the most general concepts at the top, and with more specific, less general concepts arranged below. This is motivated by Ausubel's theory that more general, superordinate concepts subsume more specific, detailed concepts [8]. Figure 1 below illustrates this approach to concept maps, and introduces many of the key themes discussed in the remainder of this document.

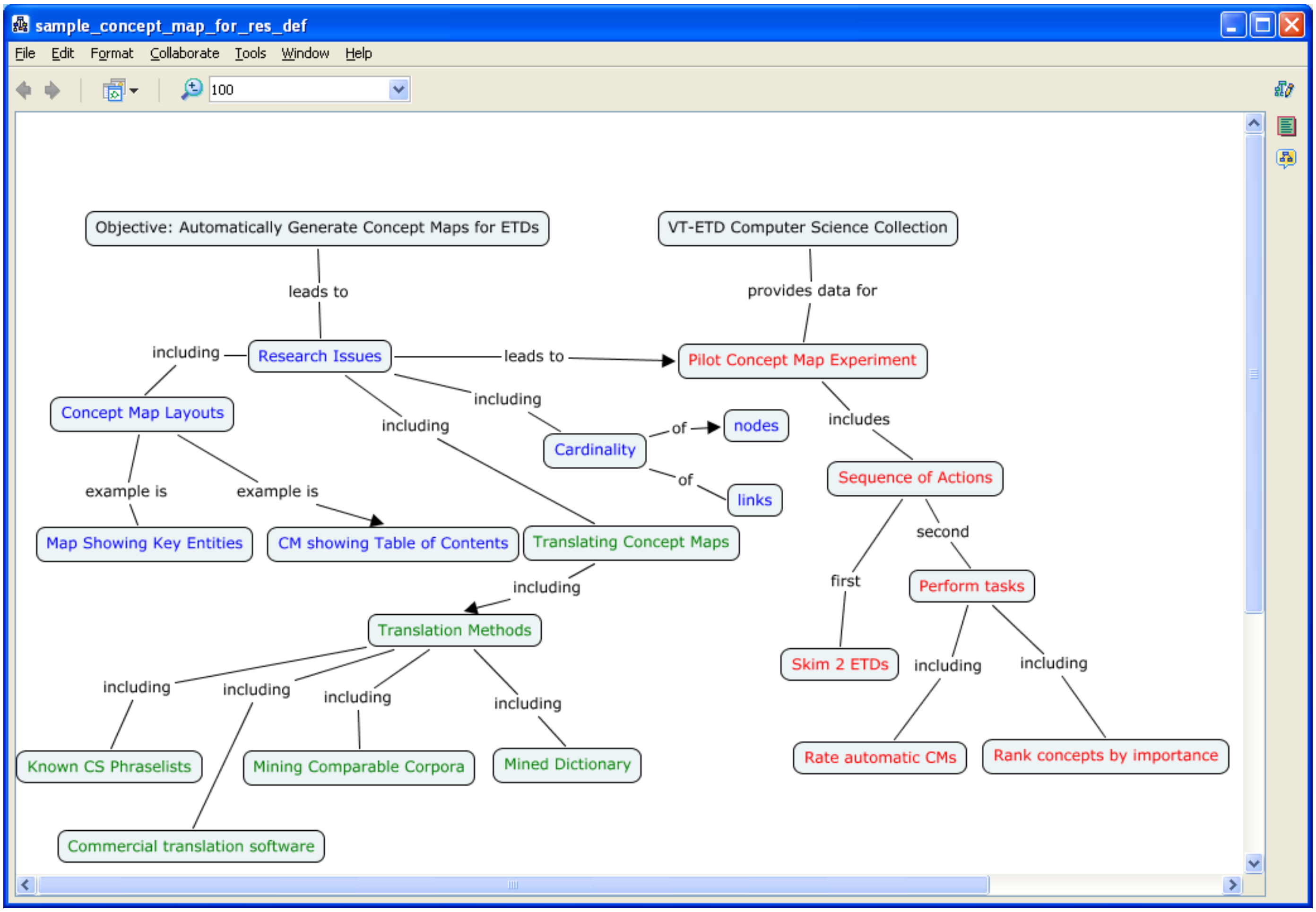


Figure 1: Sample concept map

Concept maps are themselves a special case of a semantic network [108]. Concept maps are differentiated from semantic networks in that their links can have any labels/semantics that the author wishes. Semantic networks, on the other hand, usually have a limited number of link types, with well-defined semantics [24]. Also, semantic networks do not have to be arranged hierarchically. By convention, concept maps are drawn in hierarchical fashion, either from top-to-bottom or in a star-shape, with important concepts in the middle branching out to more specific concepts around the edges. There has been much research on how concept maps allow students to acquire knowledge more quickly than with traditional teaching techniques [76, 117]. However, there has been limited research [26] on using concept maps as summaries for a paper or group of papers. Currently the two most common kinds of summaries of research papers are abstracts and keyword lists. Keyword lists are inadequate because they cannot show relationships between sets of keywords. Suppose a researcher is looking for a document where topic A relates to topic B with relationship C. Unless relationship C holds, the appearance of A and B in the keyword list does not necessarily mean that the document is relevant to the researcher. Further, there are many problems with using abstracts as the primary summarization of a document [63]. Writing good abstracts is difficult, and often the abstract mentions some topics which are only briefly discussed in the paper, and omits other topics which make up large sections of the paper. Many researchers, while technically proficient, consistently follow poor sentence organization practices, creating virtually unreadable abstracts [43]. Finally, the length and form of abstracts vary widely across subject areas and especially across languages. When it comes to automatically translating abstracts, the difficulties become even more obvious. Anyone who has ever used BabelFish (tm) to translate an abstract into their preferred language knows the poor readability and ambiguities in the resulting text [6].

1.3 Hypotheses

1) Automatically-generated concept maps can augment abstracts in helping subjects determine if a document is relevant to an information need.

2) Automatically-generated concept maps can be translated via MT well enough that they can be used as a cross-language summarization tool.

A corollary to 2) is that

2a) Concept maps can be translated adequately by translating the nodes and links, without reorganizing the rest of the map [10].

The following are some research questions that this dissertation addresses.

Question 1	Could a concept map augment a standard abstract as a summary, both for the original language, and when translated into a second language?

For Question 1, I must determine the type of 'summary' I aim to produce. Some common types of summaries are descriptive, informative, and indicative [54]. Informative summaries aim to compress almost all of the useful information in a document into a smaller package, thereby serving as a surrogate for the original document. However, given the length of a typical thesis or dissertation, often over 100 pages long, it is hard to imagine how one could produce a compact summary that could include all the relevant information that a reader might wish to know. Thus I am focusing on indicative summaries, which merely help the reader decide if she wants to acquire and read the entire document.

If the answer to Question 1 is (in some empirically-defined sense) yes, then a series of other questions comes to mind.

Question 2	Can I produce concept maps quickly and easily enough for them to be useful tools?
Question 3	Is the optimal solution to use one map, or multiple maps, with each map representing (for instance) one chapter of a large document?
Question 4	What holds with regard to questions 1-3 across languages?
Question 5	When a concept map is translated into another language, should the

	structure of the map change to express the meaning more clearly in the new language?

For my research concerning English and Spanish, which are fairly similar languages, I will assume that it is not necessary to restructure the concept maps after translating them. While I know of no research on concept maps that specifically deals with this issue, there has been research on translating a type of semantic graph called a knowledge graph. Liu [66] and Zhang [126] dealt with producing and translating knowledge graphs for both English and Chinese documents, as well as translating between the two. Their conclusion was that it was indeed possible to translate knowledge graphs between two very different languages by using a few transformations. It should be noted, however, that knowledge graphs strongly limit the number of possible relations between concepts, whereas in concept maps the allowable types of relations are theoretically unlimited.

Experience shows that drawing good concept maps is tedious and time-consuming for domain experts, and that novices will often draw concept maps with overly literal or superficial concepts [29]. Also, novices will cluster concepts for superficial reasons, such as temporal proximity, while an expert would identify a more complex relationship [74]. Essentially, then, I need a way to draw concept maps in large quantities while maintaining uniformity, which requires that they be produced automatically, and with reasonable quality.

Some of the challenges I have explored include automatic generation of concept maps, and testing the value of the resulting maps as indicative summaries. Another of the challenges of this research has been the problem of testing: How do we know that readers actually receive added benefit from looking at a concept map in addition to solely reading an abstract? How can we compare the benefits of a concept map (or set of maps) against the benefits of an abstract? In my experimental sections below I lay out the five experiments I have done to test my assertion that concept maps can enhance abstracts as summarization tools.

1.4 Outline of this dissertation

Chapter 2 gives a review of the literature regarding all of the disparate fields that this dissertation addresses. Chapter 3 gives a description of a pilot experiment I conducted using hand-drawn concept maps, used as summaries of English and Spanish research papers. Chapter 4 describes my initial attempts at automatically producing concept maps using statistical measures of term co-occurrence. Chapter 5 describes my further work in producing concept maps, using natural language processing (NLP) aware tools combined with an ontology of a specific technical field. Chapter 6 describes my work which uses a collection of Spanish ETDs as a corpus from which translations of technical phrases can be found. Chapter 7 describes my final experiment, a cross-language experiment which combined automatically-generated concept maps that I produced using techniques from Chapter 5 along with machine translation techniques that I described in Chapter 6. Chapter 8 gives my conclusions based upon my user experiments, as well as laying out several paths for future work.

2 Related work

This review will first look at how concept maps have been used as learning enhancement tools and for knowledge generation, capture, and representation. Then it will provide background on attempts to automatically generate concept maps.

2.1 Concept maps as learning enhancement tools

As envisaged by Novak, concept maps are directed acyclic graphs (DAGs), though some recent research recommends loosening the acyclic requirement [101]. Concept maps are based on the learning psychology theories of David Ausubel [8, 9]. There are many studies on the use of concept maps as learning tools. The key idea is that the learner assimilates new concepts and propositions into his or her existing conceptual and prepositional frameworks. Malone and Dekkers [70] aptly described concept maps as "windows to the mind" of students, "for seeing in (by the teacher and other students), for seeing out (by the student) and for reflecting on one's own perceptions (by everybody)". Concept maps encourage students to use meaningful-mode learning patterns [78].

Concept maps often are used as part of classroom instruction, as an ongoing evaluation of knowledge within a course. They can be useful for demonstrating the changes that occur in a student's knowledge structure as students integrate new knowledge with existing knowledge [115]. This raises the question of how concept maps should be scored for grading purposes. The original method for scoring concept maps was proposed by Novak and Gowin [87]. The system they propose assigns points for valid propositions (1 point each), levels of hierarchy (5 points each), number of branchings (1 point each), cross-links (10 points for each valid one), and specific examples (1 point each). This scoring method, which must be performed by a human, has proven time-consuming but gives a great deal of insight into the student's knowledge structure. Others tried comparing a student's map to that of an expert [100]. However, fair comparison may be difficult to automate and may require complex handling akin to human judgment. One method to score student concept maps versus expert maps with potentially varying structure was presented by Marshall et al. [71]. A simplified concept map scoring scheme was proposed by Shaka and Bitner [105], which uses Novak and Gowin's scheme [86] as a starting point, but provides simplified analysis of important map characteristics. Several map properties, including propositions, branches, hierarchies, examples, cross-links, and others – are given a rating from 0 to 4, rather than just being characterized.

There is widespread interest in various applications of concept map-style interfaces for hypermedia. These maps have been referred to as navigation maps or graphical browsers (note that this research predates the World Wide Web). Carnot et al. [21] compared a concept map-based browser to a World Wide Web page-based interface with respect to ease in finding information necessary to complete a series of search tasks. The results indicate that the concept map-based interface resulted in better search performance for all learners, although learners who sought deeper understanding benefited the most. These results were backed up by a later study by Weideman and Kritzinger [116].

Recent research has indicated that the use of concept maps in a navigational setting does not change the user's knowledge structure [85]. That research suggested the need for different maps for navigation and for learning, and the desirability of designing maps with a specific purpose in mind. Another important implication from [85] is the need for learners to be actively engaged.

Willerman and MacHarg [122] found that using concept maps as an advanced organizer produced a significant increment in students' achievement in physical science areas. Others have studied using concept maps as a tool for the collaborative construction of knowledge [98]. Esiobu et al. [33] compared concept mapping and V-diagramming as a summarization or study strategy after regular classroom instruction, versus a control group that used neither tool. Both treatment groups did better than the control group on multiple-choice achievement tests.

A group at Texas Christian University has developed something similar to concept maps, called knowledge maps. Researchers at TCU have tested the use of knowledge maps as an alternative to text presentations of numerous topics [46, 60, 93, 94]. Notable differences between concept maps and knowledge maps are as follows:

- Knowledge maps use linking phrases drawn from a limited set, and these are usually abbreviated, such as *is_a*, *part_of*, and *example*. This restriction on linking terms also can affect what appears in the nodes. Concept maps usually allow free text to be used as linking text.
- Knowledge maps are not typically restricted to a hierarchical format. Wiegmann et al. [119] explored the effects of map format and found that a map's structure has a large effect on the map's effectiveness in knowledge retention. Typically, concept maps are drawn in hierarchical fashion.
- Knowledge map nodes are sometimes concepts in the sense that Novak defined them, and sometimes they are ideas or even whole paragraphs. Novak's

definition of "concept" limits node contents to concepts, which allows for a more explicit representation of the interrelationships among concepts.

One study by the TCU group is of particular interest since it used bilingual knowledge maps (Bi-K maps) to enhance foreign language vocabulary retention [10]. In this study, the researchers produced (by hand) bilingual German-English knowledge maps by splitting each node horizontally and placing a German word in the top half, with its English translation in the bottom. Each node was connected to at least one other split node, producing a free association map. Bi-K maps use nine types of link relations: *definition*, *characteristic_of*, *part_of*, *next*, *influences*, *leads_to*, *example*, *analogous_to*, and *comment*. The researchers performed a study of 72 students unfamiliar with German, in which half were presented with Bi-K maps and the other half with a simple bilingual vocabulary list. There was also a high-semantic emphasis condition, in which the subjects were first presented with a knowledge map with the German words deleted, but with the English words and their relations intact. Then the subjects were presented with the full Bi-K map. Bi-K map learners generally out-performed list learners on free-recall, cued-recall, and multiple-choice tests. The high semantic emphasis condition did not emerge as a significant variable for enhancing performance; in fact it may have had a negative effect on recall over time. Nevertheless the Bi-K map conditions outperformed the list conditions regardless of whether they had low or high-semantic emphasis.

Concept mapping has been widely used as a knowledge elicitation (KE) tool [44, 45, 114]. Concept mapping is considered to be at least as effective as other methods, in terms of number of informative propositions identified about a domain. It has been used successfully to form mediating representations and interfaces for knowledge-based systems and tutoring systems. Concept maps clearly have a role to play in KE in the form of a simple, intuitive knowledge representation scheme [26]. In recent years, it has been recognized that the aggregate knowledge of an organization is a valuable asset that must be protected, maintained, and augmented. This has led to an increased interest in KE tools. One problem with regard to knowledge capture is making already overworked "knowledge workers" perform additional tasks. In spite of this, concept maps have been employed successfully in knowledge generation, capture, and representation [77].

Recent work by White, Song, and Liu [118] has concentrated on semi-automatic construction of concept maps to help teachers produce "stories" for classroom teaching. Their tool is an oral history search interface with two components: "retrieve" and "store/use". The test collection was of oral history recordings. Using the "retrieve" tool, searchers submit text queries and as output they receive segments of the oral history archive. These segments, as well as names of people and places, can be dragged onto the canvas interface to become part of concept maps. Users have the ability to annotate nodes, as well as to assign relationships between the nodes from a pre-defined list. The tool was used by school teachers to produce both simple and complex stories for classroom instruction. White et al. [118] found that teachers using the concept mapping tool produced stories with significantly more sources of evidence than the teachers who did not use the concept mapping tool, at least for the complex story task.

2.2 Automatic production of concept maps

The first work involving the automatic generation of concept maps was apparently done by Gaines & Shaw [38]. Their system, called "GNOSIS", produced concept maps purely based on term co-occurrence, a technique commonly used in information retrieval systems ([109], [17]). Their system was able to find sets of related terms that would have been discovered by a careful reader, but which typically differ from the list of important topics found in the paper's table of contents. The main drawback to this approach is that the nodes are labeled but the links are not.

Saito et al. [102] used a somewhat more sophisticated approach to generate concept maps of Japanese documents. They used the "Chasen" Japanese morphological analyzer to find the word boundaries in documents, and to determine the part-of-speech of each word. They used the nouns as nodes, and prepositions and conjunctions as links between these nodes. This was part of a collaborative system in which the automatically generated map became a starting point, and then students (located remotely) could suggest and make changes to it. Apparently this system has not yet been adapted for English use.

A yet more sophisticated system to automatically generate concept maps has been developed by Rajaraman and Tan [91]. Their system creates so-called "concept frames", i.e., Noun Clause-Verb Clause-Noun Clause triples. The noun and verb clauses are found by using a context-free grammar, which is derived from well-established natural language processing research. They also used synonym sets, derived from WordNet [34], to cut down on the number of concept frames. Their research showed that these concept frames were able to improve users' understanding of news stories as compared to a document browsing system.

More recently there has been related work at IBM in the biomedical domain [27]. This work concentrated on automatically finding relations in a collection, using grammatical information, and then graphically displaying the results. As with most other automatic generation techniques, very little was done to evaluate the effectiveness of this technique.

Other interesting research has been done by Lin at Drexel, with the AuthorLink and ConceptLink programs [16]. ConceptLink is an enhancement to the National Library of Medicine's PubMed search service. It uses the Unified Medical Language System (UMLS) co-occurrence database to suggest query terms to users. It draws what are essentially self-organizing maps (SOMs) [58]. It differs from my work in that it does not produce link labels. Also it only deals with English documents, and is restricted to the medical domain.

There have been many tools, such as Starlight [95, 97] and NSPIRE [84], which attempt to extract keywords from documents and display the keywords graphically, often clustering similar keywords together. Although related to automatic concept map generation, these tools do not attempt to find semantic relationships between the clusters, so they do not produce true concept maps.

There also has been work on visualizing association rules directly, such as INFOVIS [123] and MIRAGE [125]. This work tries to use visualizations to elucidate patterns in large datasets. Since this work is applied to arbitrary datasets, not English text, it is not finding associations between concepts as such. Also it does not attempt to label the relationship between the concepts. Thus the visualizations that are produced have little to do with concept maps. For instance, if given a sample of English text as data, it might find a relation between the terms ‘tree’ and ‘trees’, which are two forms of the same English word, and are therefore the same concept.

In summary, there has been work on automatically generating concept maps since the mid 1990's. A common issue has been that evaluating different techniques is difficult. Not even a human expert can say for certain what a 'correct' map should look like. Indeed, what the 'correct' map would be depends on the use for which the map is intended. Finally, none of the prior work addresses how well a concept map can supplement, or replace, an abstract, as a summary for a large document (especially in a cross-language setting).

To the best of my knowledge, my work is the first to be published that uses deep natural language processing (via Relex), and an ontology, to automatically produce concept maps. It is the first to attempt to translate automatically-generated concept maps into another language. It is also the first to evaluate the concept maps produced with respect to their fitness in assisting users to make relevance judgments.

2.3 Extracting concepts from free text

Entity searching, recognition, and extraction has been a topic of research since the late 1980's [73]. The first research in the field of Named Entity Recognition (NER) concentrated on recognizing the names of people, organizations [92], and geographic locations in text documents [55]. A frequently studied kind of source text is newspaper articles, since they are widely available and are presented in a fairly regular format. These were a key source of information for the Message Understanding Conferences (e.g., [113]). There also has been work on generalized keyphrase extraction tools, in which the user trains the phrase extractor on a small training set, and then uses the tool to extract keyphrases from a much larger training set. An example of this is the University of Waikato's Kea [52, 90].

However, entity-extraction is complicated by the use of synonyms, polysemy (one word having multiple meanings), lemmas (words with the same root but different morphology), and quasi-synonyms [67]. Mattox et al. conclude that semantic knowledge about a specific domain, such as an ontology, is necessary for entity-extraction, and producing such resources is non-trivial [72].

More recently, there has been a great deal of research in NER in the bioinformatic and medical domains. Most involves extracting names of genes, proteins, molecular pathways, drugs, etc., as well as relations from medical articles. Examples are ABNER [103], BioRAT [28], EDGAR [96], GENIES [35], POSBIOTM [51], and Textpresso [79]. Another tool, MedLEE [36], is designed to extract information from textual patient reports.

Research on entity extraction in other scientific fields is not as advanced as in the biomedical field. For computing, a starting point for entity extraction would be a comprehensive ontology of the entire field. One project to produce an ontology of computing was led by Krassen Stefanov at the U. of Sofia. The idea was to use the ACM Computing Curricula 2001 as a starting point, and make the necessary changes and extensions to produce an usable ontology [111]. A better approach was suggested by the ACM Joint Task Force on Computing Curriculum, which began a project to produce such a computing ontology (called the Computing Ontology Project [104]) as part of its efforts to produce a new version of its Computing Curriculum (called CC2004). This work is ongoing [22].

2.4 Use of bilingual corpora for phrase translation

The problem of finding translations of technical phrases and jargon has long been recognized. Leeds and Nakhimovsky (1979) note that "Published translations of such collocations are not readily available, even for languages such as French and English, despite the fact that collocations have been recognized as one of the main obstacles to second language acquisition" [62]. Since bilingual dictionaries typically only provide translations for words, and have limited coverage of fixed phrases, there have been attempts to use other techniques. The idea of using parallel corpora to assist machine translation dates back to at least 1984, when [81] suggested using parallel corpora to translate between English and Japanese. These ideas were furthered developed in [15] and [82]. The first large-scale use of this idea was in 1990, when IBM used the Hansard corpus (French/English) as the basis for experiments in corpus alignment [56]. Using parallel corpora to provide translations of noun phrases has been done since Fung's Chinese/English work in 1995 [37]. His technique relied on aligning large chunks of the corpus (averaging about 300 words) and finding proper nouns and noun phrases in both languages in these chunks. This assumes that one can find a sufficiently large parallel corpus in the required domain. About the same time Wu [124] identified noun phrases in an English/Chinese parallel corpus without using language-specific monolingual grammars for either language, by using an inverse transduction grammar to model both languages simultaneously. Hull [49] studied the practical problems with finding phrase translations from parallel corpora, such as the fact that the need for several iterations made the algorithms less efficient. The relative success enjoyed by these approaches has meant that for about the last ten years, the MT field has become increasingly focused on the issue of finding translation knowledge from aligned parallel corpora.[14]

Of course this approach will not work for comparable corpora, where there is no expectation that parts of the corpus beyond the word/phrase level can be aligned. Early work on comparable corpora focused on assisting translations of individual words, by using the corpus to perform word-sense disambiguation (WSD). For example Kaji and Morimoto [53] used a comparable corpus of English and Japanese newspaper articles to disambiguate the meaning of 60 English words. Current research relies on the assumption that if two words are mutual translations, then words that frequently collocate with them are likely to be mutual translations as well. This approach requires a bilingual dictionary, either a general or specialized one, as an initial seed to relate words across languages. Between related languages such as Spanish and English, this requirement can be circumvented by automatically building a seed lexicon using spelling and cognate clues [57].

There also has been limited work on using comparable corpora to extract bilingual lexicons [31]. Thus, Dejean et al. used a bilingual thesaurus provided by UMLS/MeSH for English/German to extract terms and phrases in the medical domain. Their work focused on parallel corpora, but could be extended to comparable corpora. On parallel corpora the Dejean method achieved 89.8% precision and 81.0% recall. Later work by Gaussier, Dejean, et al. developed a modified version of probabilistic latent semantic analysis (PLSA) to correlate terms using a small bilingual dictionary and a comparable corpus [41]. Their results were only marginally better than standard approaches for term extraction from comparable corpora, and were more computationally intensive than standard approaches.

Lopez-Osteñero used a standard bilingual dictionary and a comparable corpus in English/Spanish to translate noun phrases, which were then used to assist users in cross-language search tasks [68]. My methods in this area are based on their approach.

2.5 Automatic text summarization

Attempts at automatic text summarization date back to the 1950s. They have shown that human-quality text summarization is difficult since it requires discourse understanding, abstraction, and language generation [110]. While an abstract summary may more closely match readers' intuitive notions of a summary than would an extractive summary, their creation involves greater complexity and difficulty [48]. Unlike producing text extracts, producing abstracts requires several stages, such as topic fusion and text generation. Also, by using sentences from the text, an extractive summary remains closer to the original document, thus limiting potential bias from the automatic summarization program (and its authors) [69].

Thus most modern systems have adopted a simplified approach, which involves identifying spans of related text in the document, and extracting sentences representative of each text-span. In this context summarization can be defined as the selection of a subset of the document sentences which is representative of its content [5]. This is done by ranking document sentences based on relevance to a particular topic, and also by redundancy based on considering other sentences in the document. The selected sentences are the ones with the highest relevance and lowest redundancy. This is often computed using a Maximal Marginal Relevance (MMR) measure, first used in this context by Carbonell and Goldstein [20]. The text-spans are usually sentences, but paragraphs also have been considered [112]. While the quality of an abstract summary might not be as good, from the reader's perspective, as that of an extractive summary, it is often good enough for the reader to understand the main ideas of the document [5]. However, McDonald and Chen [75] have noted that even with indicative summaries, poor readability can detract from information content.

2.6 Cross-language summarization

There has been less research in producing cross-language summaries than there has been in producing monolingual summaries. A workshop at SIGIR 2002 dealt with cross-language summarization and laid out a broad plan for future research [42]. The dominant paradigm in text summarization, which is text-span extraction, leads to readability problems when put through an automatic translator, since the MT system has to translate sentences that were pulled from disparate parts of a document. Thus it has all the problems of standard MT, and more. This has hampered much work on automatic cross-language summarization.

3 Feasibility of using concept maps to supplement abstracts

3.1 Experiment 1

In order to test the hypothesis that concept maps can be a better summary than an abstract, both for one language and across languages, I conducted an initial experiment. Before trying to develop complicated techniques to automatically produce concept maps, it was prudent first to test if hand-drawn concept maps, by expert users, could help students identify relevant papers, even if those students were not experts in the paper's specific subject matter. I wanted to test this for both one language (English) and across languages (English and Spanish). I selected 8 short research papers, four in English and four in Spanish. The four English papers were on the topic of digital libraries, and the four Spanish papers concerned distance learning. A graduate student, Rao Shen, skilled at drawing concept maps, drew concept maps for the English papers. Another graduate student, who is a native Spanish speaker (Jesus Trespalacios), drew concept maps for the Spanish papers, after some training about concept maps. The concept maps were drawn based on the full text of the papers, so they might contain information that was not in the abstract.

I tested 28 subjects in total, all proficient in English, with no knowledge of Spanish. Subjects never saw the full documents, only the abstracts and concept maps. The subjects were divided into two groups, with 14 in each group (n=14). Group 1 was only given abstracts, while Group 2 was given both abstracts and concept maps. Subjects were given 4 questions, one at a time, and asked to rank the 4 papers (either English or Spanish) from most to least relevant for each question. Questions 1 and 2 dealt with the 4

English papers, and Questions 3 and 4 dealt with the 4 Spanish papers. The four questions were as follows:

Table 1: Relevance questions in Experiment 1.

Q1	Which paper is about the development of a fully functional virtual union catalog?
Q2	Which paper discusses the "mediator and wrapper" model of building digital libraries?
Q3	Which paper is about the advantages and disadvantages of distance learning?
Q4	Which paper discusses dividing a system into a logical module and an interface module?

Below are the concept maps used in Experiment 1. The same 4 concept maps were used for the first 2 questions, and the other 4 concept maps (translated from Spanish by Jesus Trespalacios) were used for questions 3 and 4.

Table 2: Figures for each questions in Experiment 1

Questions	**Concept maps shown in figures following**
Q1, Q2	Figures 2,3,4,5
Q3, Q4	Figures 6,7,8,9

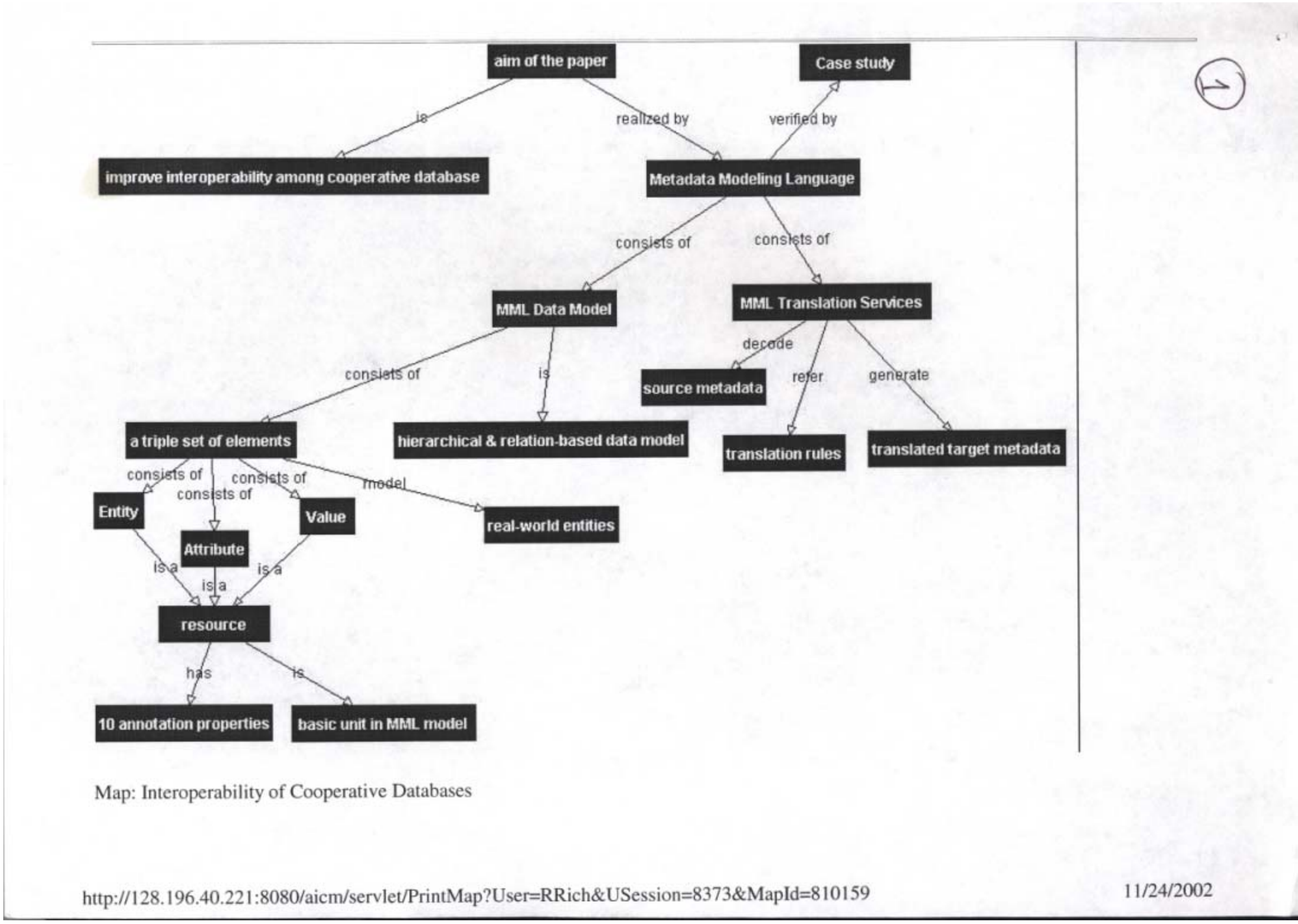
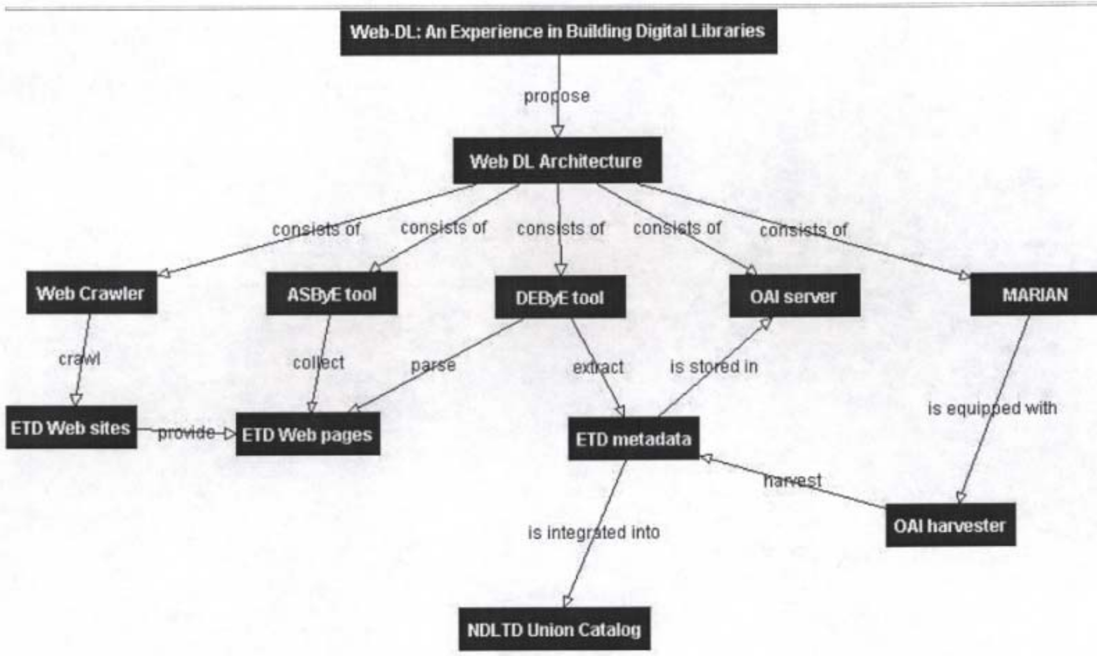


Figure 2: Interoperability of Cooperative Databases with Metadata

Map now avaible for printing Page 1 of

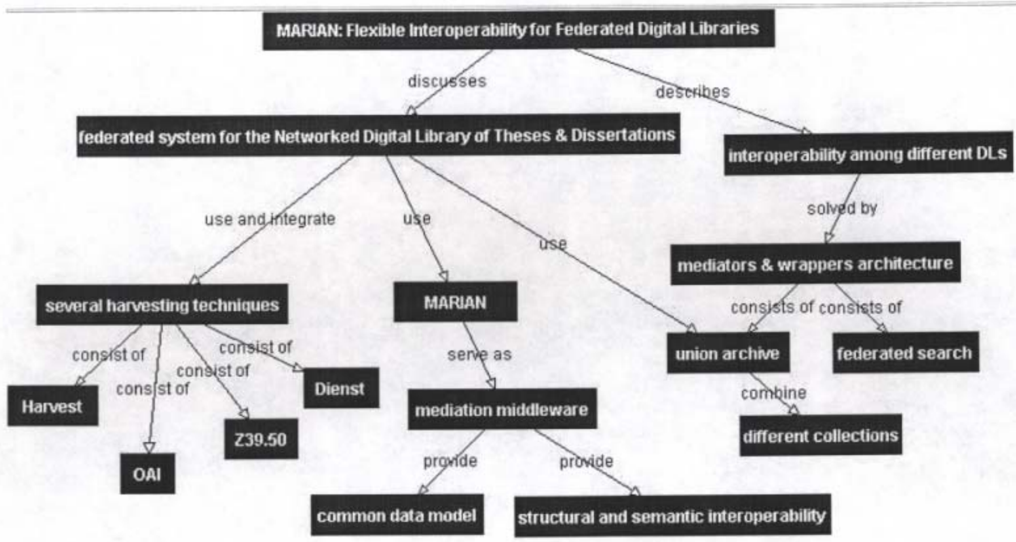


Map: Web-DL

Figure 3: Web-DL: An Experience in Building Digital Libraries from the Web

Map now avaible for printing

Page 1 of 1



Map: MARIAN Flexible Interoperability for Federated DLs

Figure 4: MARIAN: Flexible Interoperability for Federated DLs

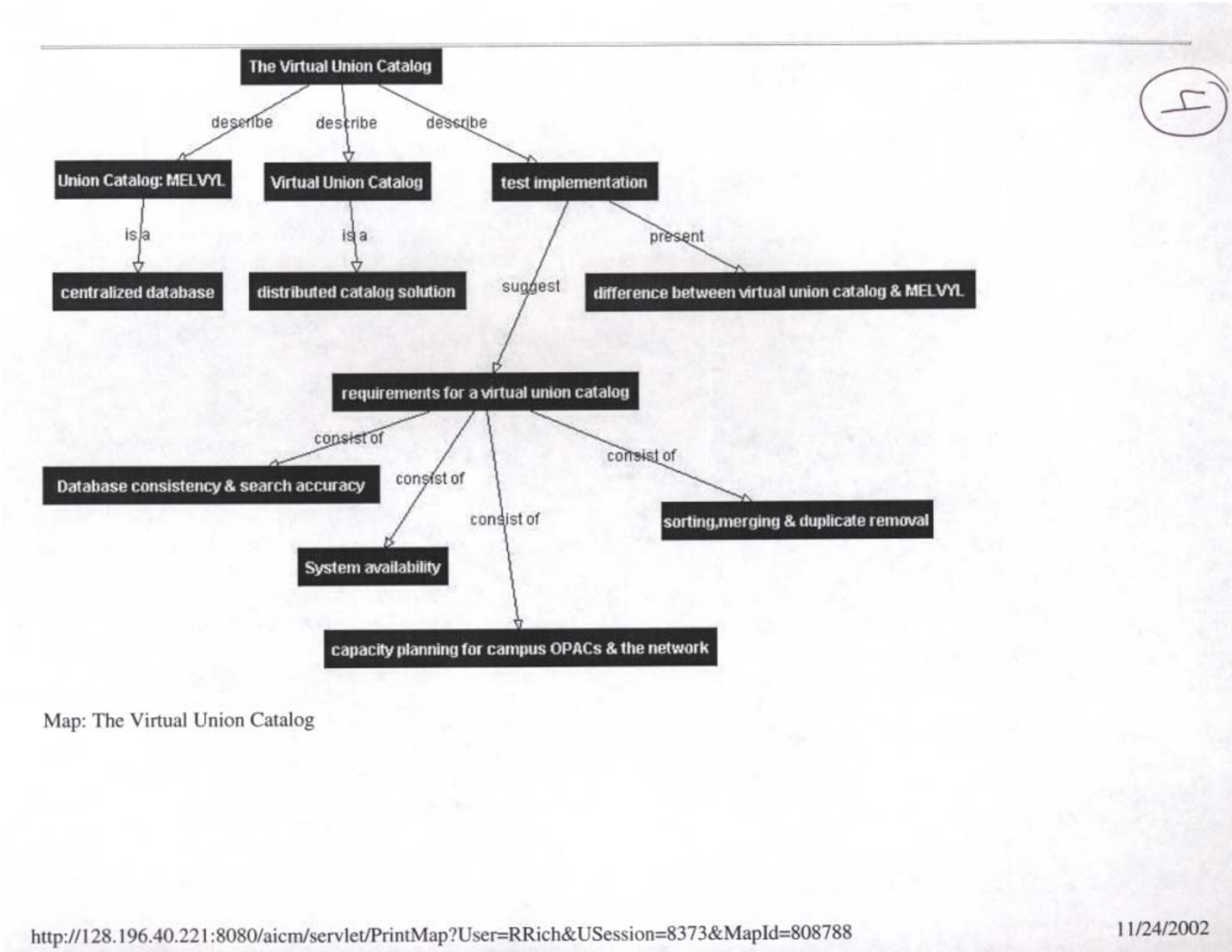


Figure 5: The Virtual Union Catalog

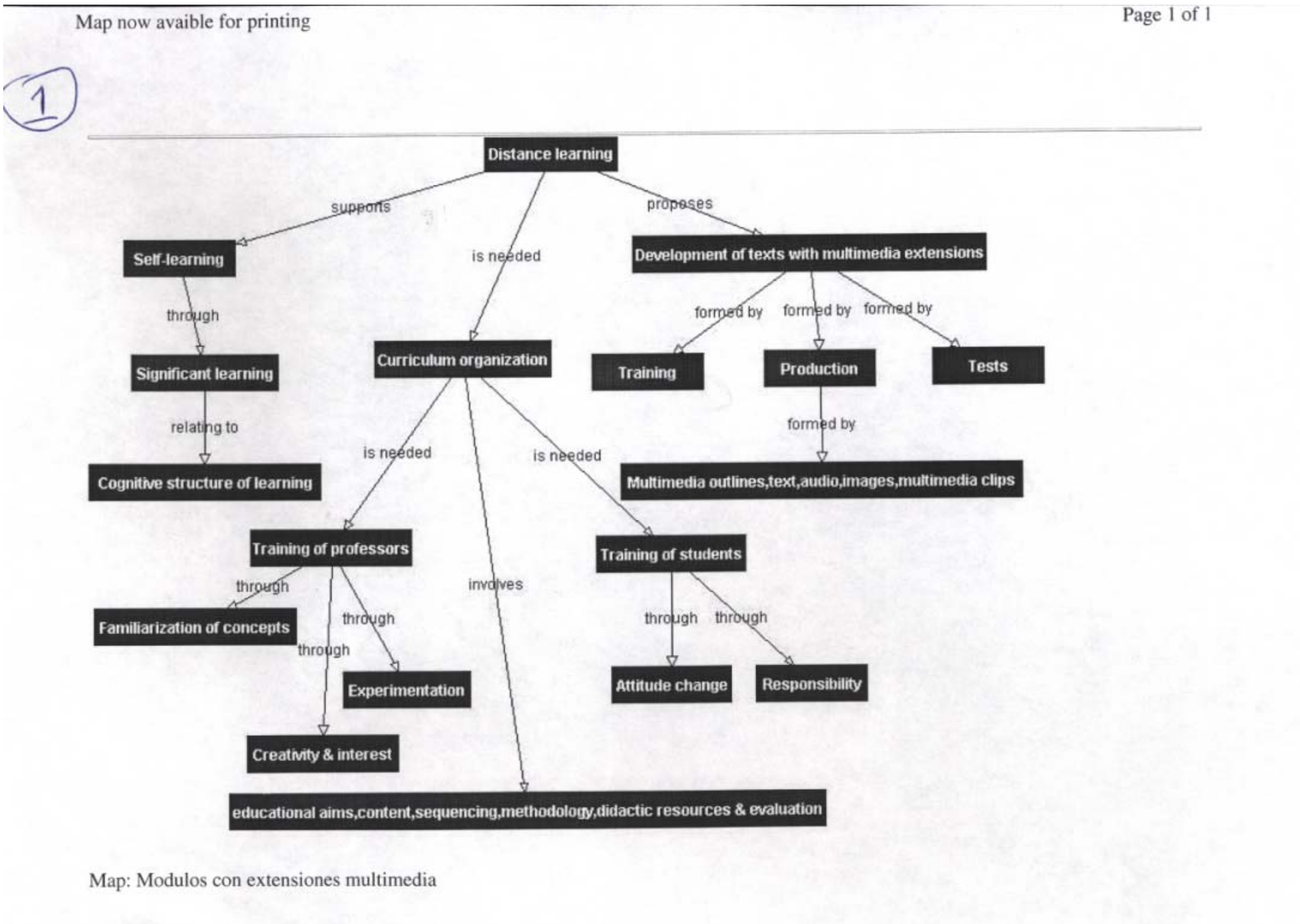


Figure 6: Map of "Modulos con extensiones multimedia para apoyar la educacion a distancia" (Modules with multimedia extensions)

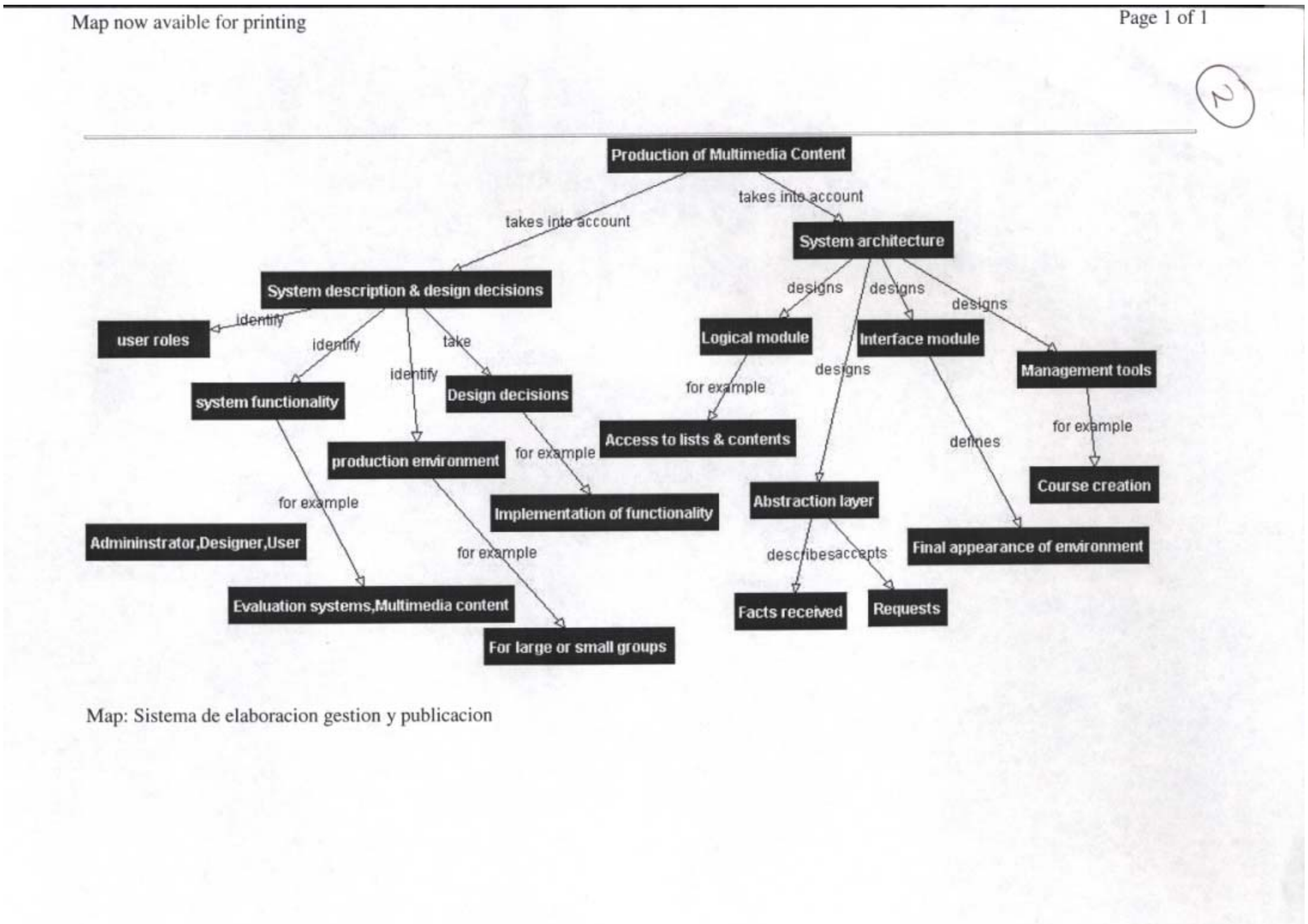


Figure 7: Map of "Sistema de elaboracion, gestion y publicacion de contenidos multimedia" (Elaboration, Management, and Publication System for Multimedia Content)

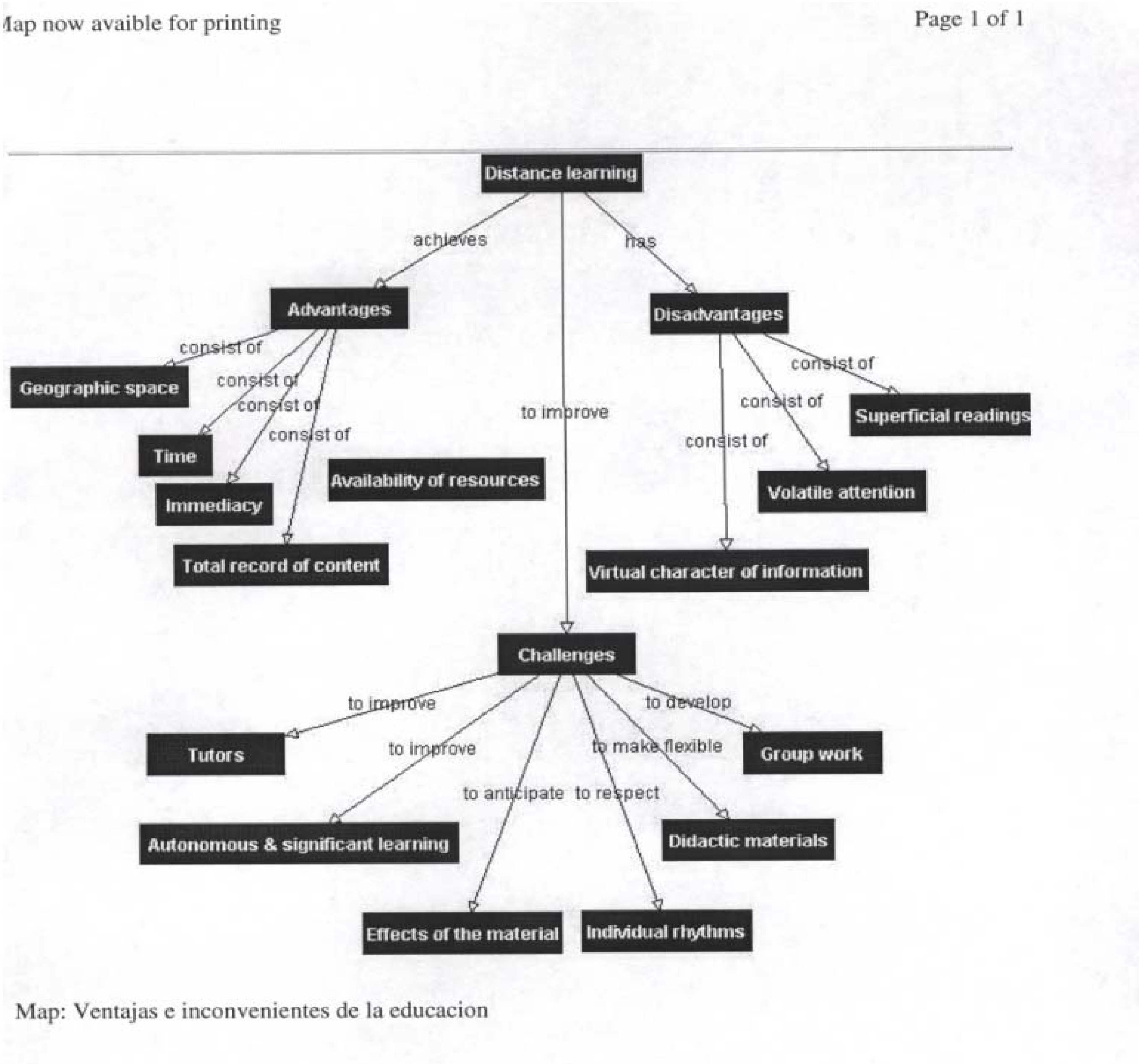


Figure 8: Map of "Ventajas e inconvenientes de la educacion a distancia a traves de Internet" (Advantages and Disadvantages of Distance Learning via the Internet)

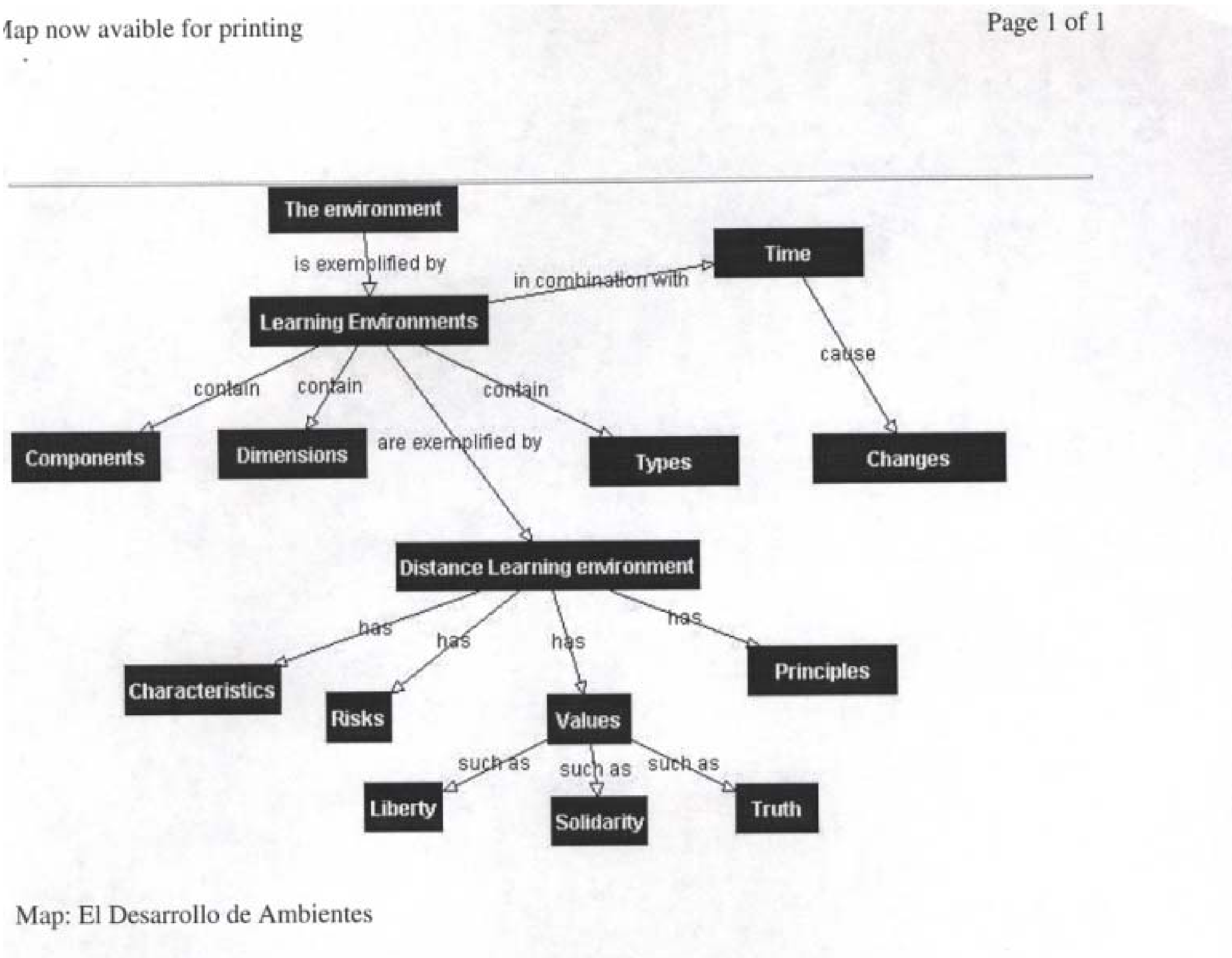


Figure 9: Map of "El desarrollo de ambientes de aprendizaje a distancia" (The Development of Distance Learning Environments)

For the Spanish papers, the subjects were given the original Spanish language abstract and a machine translation (provided by Altavista's BabelFish Translation Service) [59], [61]. The subjects in Group 2 also were given concept maps of the original documents, with the nodes and links each separately translated into English. In order to avoid learning and/or fatigue effects, half of each group was first given the English papers, and the other half was first given the Spanish papers. The amount of time the subject took to answer each question was recorded using a stopwatch.

After a subject had answered all four questions, he/she was asked to rate the helpfulness of the abstracts (and concept maps if in Group 2) on a five-point Likert scale. Those in Group 1 answered 3 questions. Subjects in Group 2 were given those same questions, plus two additional questions concerning the helpfulness of the concept maps.

3.2 Experimental results

For groups 1 and 2, the amount of time spent answering each question was recorded. The time taken for each question, and the total time, was roughly the same for both groups. In particular, the differences were not enough to reject the assumption that the means are equal ($p=0.05$).

I compared the rankings of the papers' relevance against the expert's ranking. The relevance rankings are found in Table 3. The expert ranking for the English papers was done by the researcher, and for the Spanish papers was done by Marcos Gonçalves, who can read Spanish competently. I used a simple N-dimensional Euclidean distance measure to determine how different the subjects' answers were from the experts' answers. Using this measure, I found that for questions 1, 2, and 3, there was no significant difference between the Group 1 and Group 2 students. This could be because the questions were too simple, so the students did not need much detailed information to answer them. Rajaraman and Tan [91] found similar results when the questions were too easy given the information that the subjects had access to. In question 4, which was a question about the Spanish papers, the Group 2 students answered significantly better (p=0.03, see). The subjects' answers are given below. By way of example, the first user in Group 1 thought that paper 2 was the most relevant, followed by paper 1, then paper 3 – and that paper 4 was the least relevant.

Table 3: Rankings of paper relevance from question 4. User rankings which agreed with the expert ranking are given in boldface.

Group 1	Group 2
2,1,3,4	**2,1,4,3**
2,3,1,4	2,4,1,3
2,1,3,4	2,1,3,4
2,4,3,1	2,1,3,4
2,1,3,4	2,1,3,4
2,1,3,4	2,3,1,4
3,2,1,4	**2,1,4,3**
2,1,4,3	2,1,3,4
2,4,1,3	**2,1,4,3**
2,1,3,4	2,4,1,3
2,3,1,4	2,1,3,4
2,1,3,4	2,1,3,4
2,1,3,4	**2,1,4,3**
2,1,3,4	**2,1,4,3**

These listings in order of relevance of the papers had to be transformed into rankings for each paper, since one cannot defensibly do arithmetic on the paper numbers, which were assigned arbitrarily by me before the experiment. For example, the listings in order of relevance (2,4,1,3) is transformed into the paper ranking (3,1,4,2), since the first paper was ranked third, the second paper first, the third paper fourth, and the fourth paper second. After all the listings have been transformed in this way, one can perform calculations on the rankings. A single value summary of the correctness of each list was calculated by computing the distance between the expert's ranking and each subject's ranking, using the following Euclidean distance formula:

$$d = \sqrt{\left(r_{e_1} - r_{s_1}\right)^2 + \left(r_{e_2} - r_{s_2}\right)^2 + \left(r_{e_3} - r_{s_3}\right)^2 + \left(r_{e_4} - r_{s_4}\right)^2}$$

where r_{e1} is the expert's ranking of the first paper and r_{s1} is the subject's ranking of the first paper, and so on. The result of the t-test for question 4 is given in below. The result is significant (*p=0.05*).

The students in both groups were asked to rate the perceived summarization effectiveness of the abstracts and concept maps on a 5-point Likert scale. For Group 2, on the English papers, the participants reported that the concept maps were significantly more helpful than the abstracts (*p=0.05*). For Group 2, on the Spanish papers, the participants reported that the concept maps were significantly more helpful than the abstracts (*p=0.05*). See Table 2 for a summary of the results. The results that are significant at the p=0.05 level are highlighted with an asterisk (*). For the first four questions, lower scores mean that the subjects answered closer to the expert's answer. For the two Likert questions, the first number is the average rating of the abstracts (given by the subjects), and the second number is the average rating of the concept maps (higher number means more helpful). is a summary of the experimental results for all questions.

Table 4: Results for experiment 1 (lower is better)

	Group 1 average	**Group 2 average**	**p-value**
Q1 (Eng.)	1.66	1.38	0.527
Q2 (Eng.)	1.66	1.13	0.185
Q3 (Span.)	1.71	1.10	0.209

Q4(Span.)	1.68	0.98	0.030 *
Likert (En)	N/A	3.6, 4.4	0.022 *
Likert (Sp)	N/A	2.7, 4.3	0.001 *

3.3 Discussion of results

For 1 of the 2 questions concerning the Spanish papers, subjects in Group 2 produced lists that were significantly closer to the expert's list. It could be that the ranking task was too easy to bring out differences, or that four papers was not enough to make ranking difficult. The second Spanish task did produce a significant result, as well as the user-perceived helpfulness as judged by a Likert scale. This set of results was sufficiently encouraging to warrant further research.

4 Automatic generation of concept maps

4.1 Language independent architecture

My initial efforts focused on using tools that are language dependent, and also modular. Essentially this means that I can unplug the code needed for handling, say, Spanish, and replace it with code to handle a different language, and the results will still make sense. My conjecture was that I need only three modules to add a new European language. They are the following.

1. Stemmer

2. Noun-phraser.

3. Knowledge of relative frequencies of words in this language (stopwords, etc.).

The stemming algorithm depends on the particular language, or family of languages, one is dealing with. The best choice of stemmer or noun phraser varies based on the language. For instance, simple Porter stemmers perform poorly on Romance languages [47]. For agglutinative languages such as German, one noun may function as a noun phrase, alleviating the need for a complex noun phraser.

Unfortunately, the quality of concept maps (hereafter referred to as maps for brevity) produced this way was not as good as I expected. To enhance the quality I decided that I needed to use state of the art natural language processing (NLP) tools, which are tuned for one specific language. This is discussed further in Chapter 5.

4.2 Choice of document genre for tests

Since a review of the literature reveals very little work on summarizing large documents, I selected this as the first domain for further tests for my ideas about concept maps. Due to the work of the Networked Digital Library of Theses and Dissertations (NDLTD) [83], electronic theses and dissertations (ETDs) have become a readily available genre of large documents. They also can be found in languages other than English. Thus I use them as the test genre for my automatic concept map generation.

4.3 Design layouts for concept maps

According to the work of Wiegmann et al. [119], it is important to be cognizant of the importance of map structure on knowledge transfer. To ensure map effectiveness, I planned an experiment to test which of three concept map "design layouts" are preferred by users. The choices are Table of Contents-based maps, whole-document maps, and chapter-by-chapter maps. Details of the experiment appear in the next sections.

4.3.1 Whole document maps

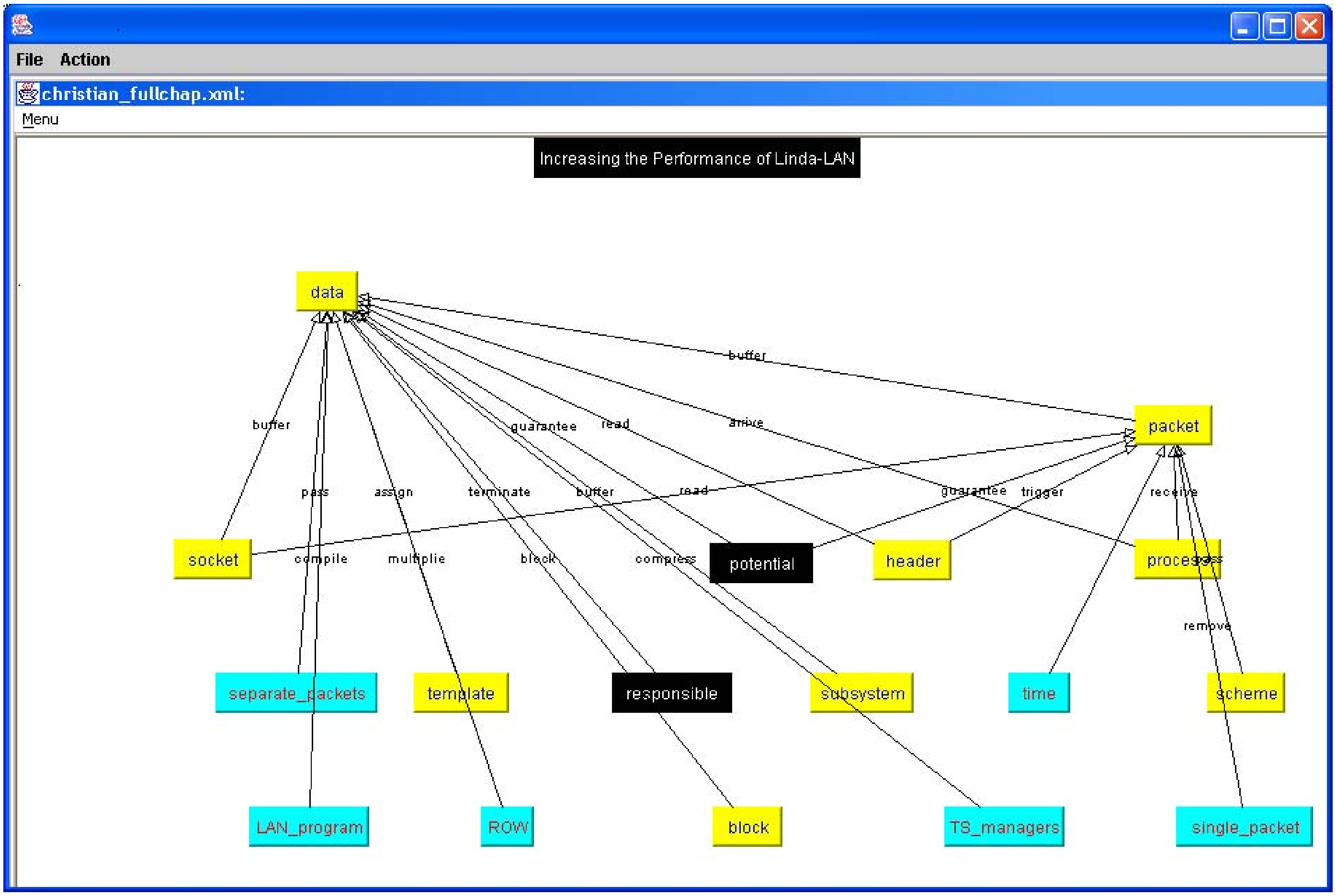


Figure 11: Whole document map of an ETD. Information obtained from "Improving the Performance of LINDA-LAN" by Jason L. Christian. Top node is the dissertation title, yellow nodes are nouns, cyan nodes are noun phrases identified found by noun phraser, and black nodes are adjectives.

Following is the algorithm I used to produce whole document maps, as in Figure 11.

Algorithm 1:

1. Run part-of-speech tagger on whole document. I use the MontyTagger program developed by Hugo Liu at M.I.T. [65], released under the GPL. It claims 95% accuracy on typical U.S. English non-fiction text, comparable to a Brill tagger.

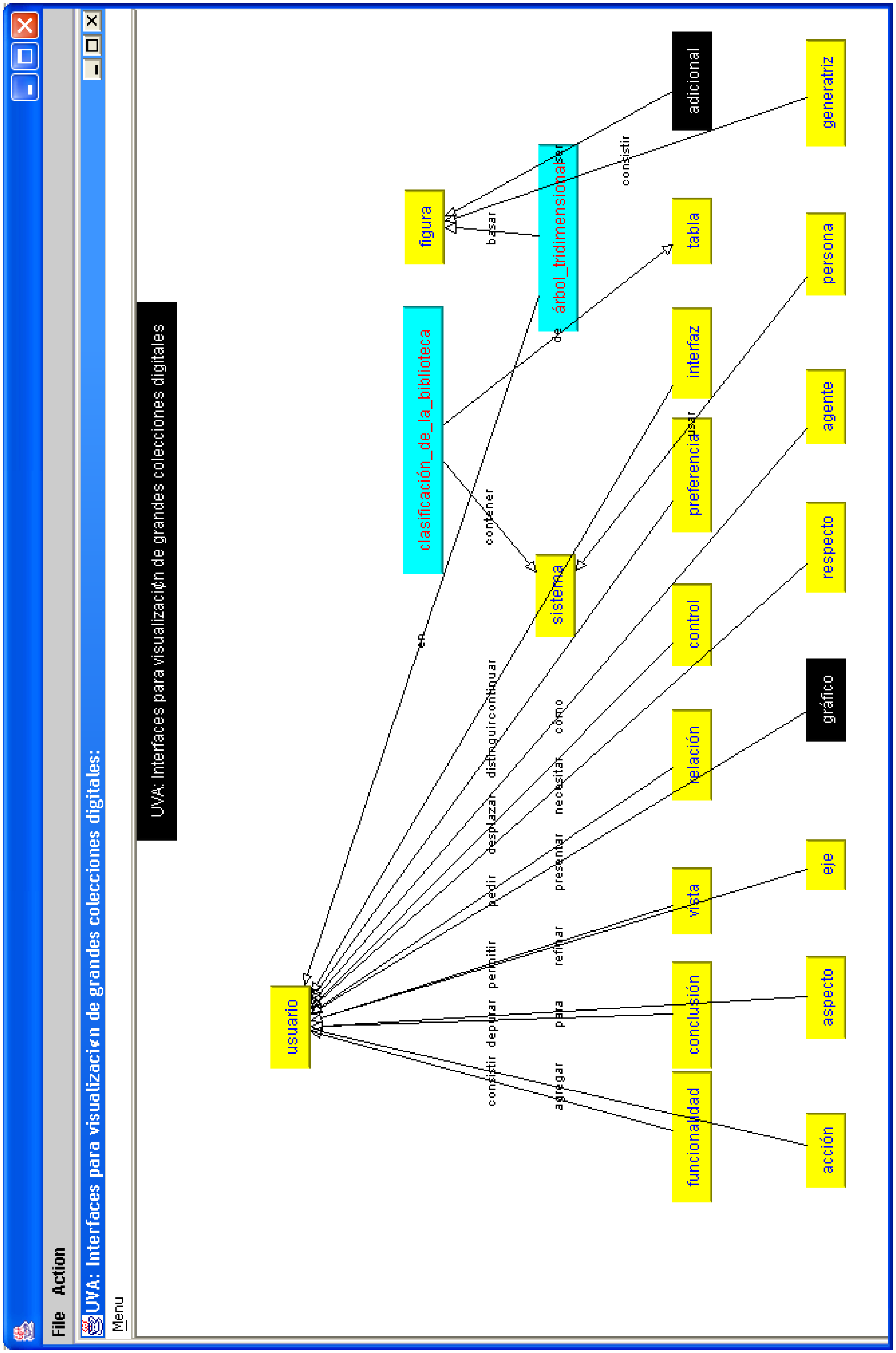
2. Find association rules consisting of words/noun phrases in the document. For this I use the highly-optimized Apriori package developed by Christian Borgelt [12, 13]. The relations which are found become the nodes. These nodes are color-coded based on the part-of-speech found in step 1, shown in Table 5:

Table 5: Mapping of parts-of-speech to colors for Whole document maps (Figure 11, Figure 12).

Part-of-Speech	**Color**
Noun	Yellow
Noun phrase	Cyan
Adjective	Black

These colors were chosen for maximum contrast with the node text.

4. For each relation, look for the connecting words (verbs, conjunctions, and prepositions) that occur between the two related words in the text. Select the most frequently occurring word, taking into account the relative frequency of the word in the language. Take for instance the relation "data" -> "socket". In these maps, the directionality of the arrow simply implied that the term "data" often co-occurred with the term "socket", but not necessarily in that order. For the relation "data"-> "socket", the most common connecting word is the preposition "for". However, since "for" is such a common word, my code selects the verb "buffer" instead, based on its frequency in the document versus its rarity in English in general. An example sentence would be "Data is buffered for use in sockets." Since the term "buffer" (which can be a noun or verb) may not actually be functioning as a verb in its co-occurrences with "data" and "buffer", this approach does not always give good results. More advanced techniques for finding labels for the relationships are discussed in Chapter 5.



File
Action
Menu
UVA: Interfaces para visualización de grandes colecciones digitales
usuario
funcionalidad
conclusión
vista
relación
control
preferencia
interfaz
tabla
acción
aspecto
eje
gráfico
respecto
agente
persona
generatriz
adicional
sistema
clasificación_de_la_biblioteca
figura
árbol_tridimensional
consistir
depurar
permitir
pedir
desplazar
distinguir
continuar
agregar
para
refinar
presentar
necesitar
como
en
contener
basar
de

Figure 12: Previous page: Whole Document map of a Spanish ETD. Information obtained from the dissertation "UVA: Interfaces for visualization of large digital collections" (English translation) by Carlos Proal Aguilar.

4.3.2 Chapter level maps

One possible improvement to whole document maps is to treat each chapter as its own document and make a map for each one. Then I can display all the maps together on one page, with the chapter 1 map on top, and the chapter 2 map below that, etc. This requires the user to scroll down to see all the chapter maps.

Here is the algorithm:

Algorithm 2: Chapter Level Maps

1. Split document up into chapters, and apply **Algorithm 1** to each chapter.

2. Add the new chapter map beneath the previous chapter.

(Optional) Draw links when the same term appears in different chapters.

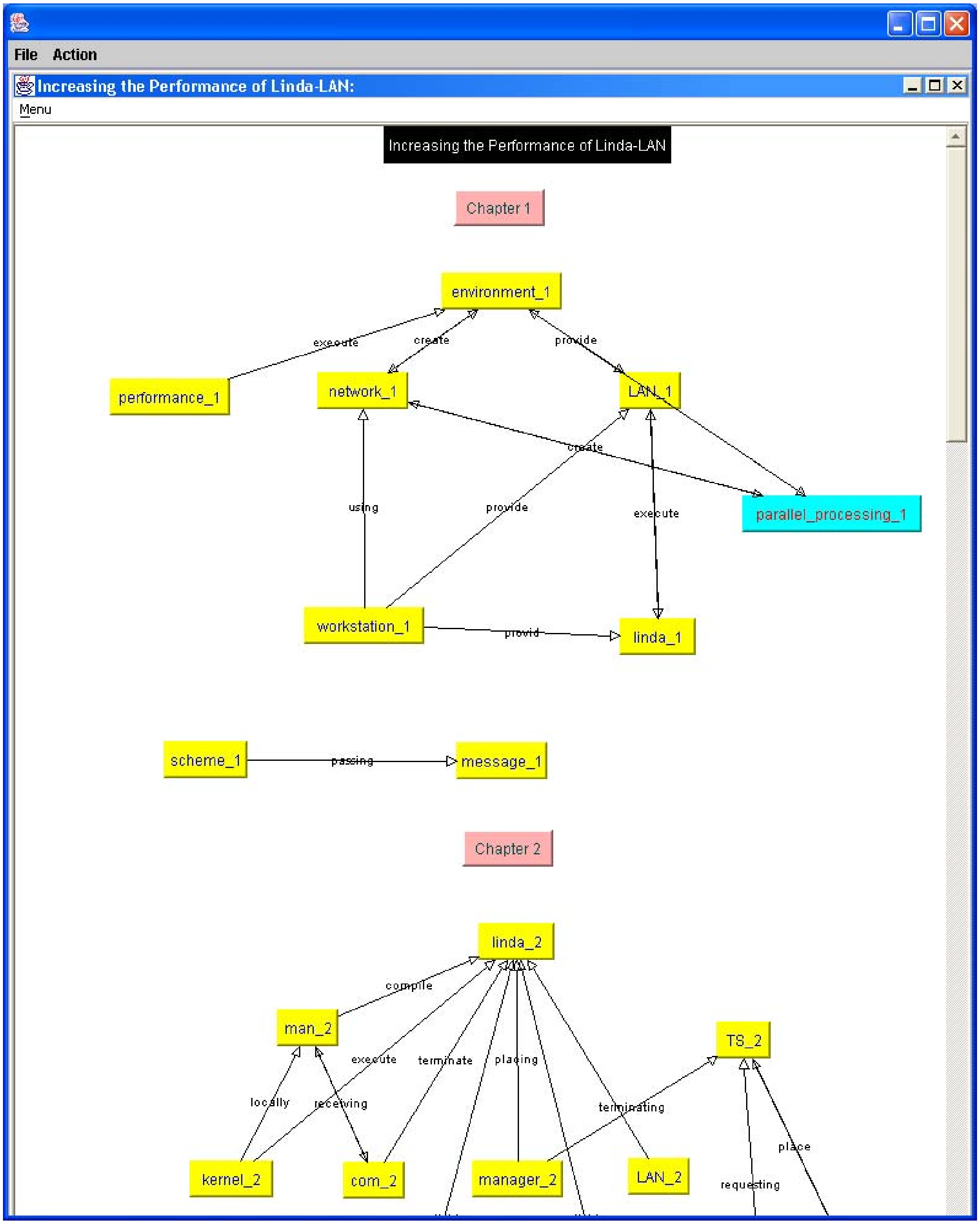


Figure 13: Chapter Level Map of a dissertation. Information obtained from "Increasing the Performance of LINDA-LAN" by J. L. Christian. Top node (black) is the dissertation title, pink nodes give chapter numbers, yellow are nouns, and cyan are noun phrases.

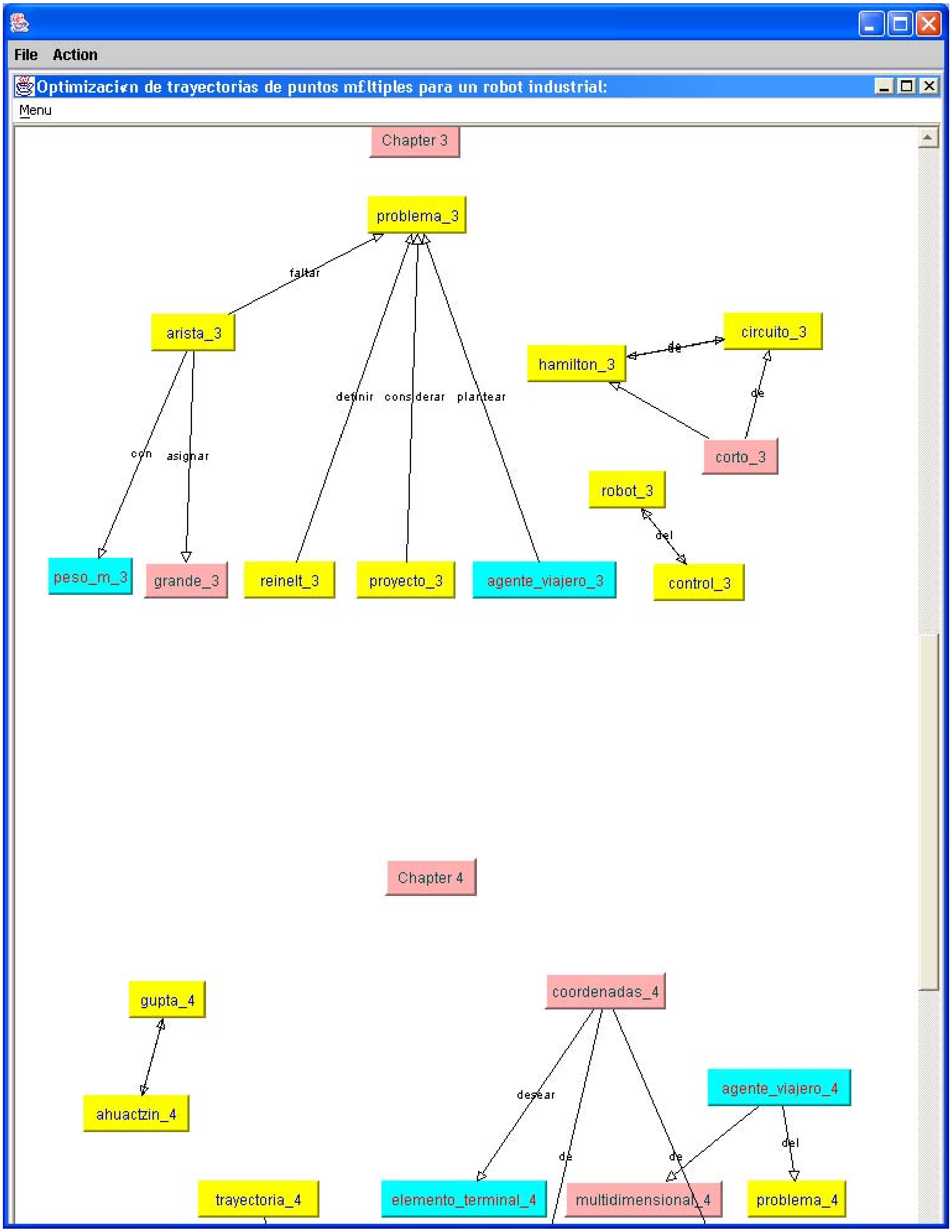


Figure 14: Chapter Level Map of a Spanish thesis. Information obtained from "Optimization of multiple-point trajectories for an industrial robot" (English translation) by Ruth Lezama Morales. As a result of scrolling down, this shows the concept maps of chapters 3 and 4. Pink nodes are chapter numbers, yellow are nouns, and cyan are noun phrases.

4.3.3 Table of contents maps

During my first pilot experiment, I noticed that the concept maps which our Spanish expert drew often used the table of contents as the skeleton of the map, and then simply elaborated on it. Since producing maps in this manner can easily be automated, this motivated the creation of Table of Contents maps.

Algorithm 3: Table of Contents Based Maps

1. Extract chapter and section heading text from table of contents, draw these as the 1st two levels of nodes and links (colored blue and red).
2. Look for words in the chapter that correlate with words in section headings. Add these as a third level of nodes + links (colored green).
3. Look for other words that correlate with words in the 3rd level, and make then into a 4th level (colored yellow)

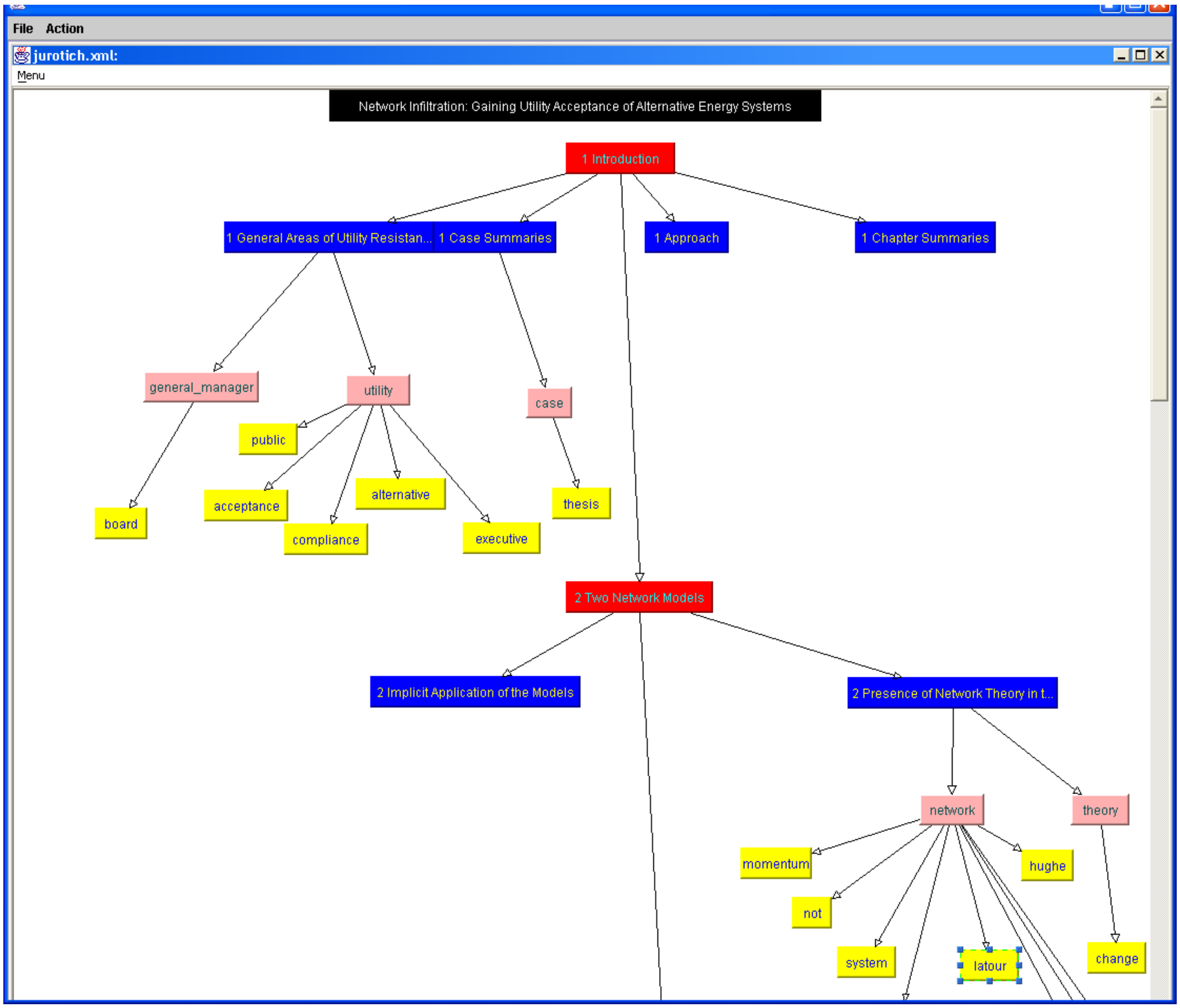


Figure 15: Table of Contents map for a master's thesis. Information obtained from "Network Infiltration: Gaining Utility Acceptance of Alternative Energy Systems" by Theresa Jurotich. Black node is the thesis title, red nodes are chapter titles, blue nodes are section headings, pink nodes are words that co-occur with words in the section headings, and yellow nodes are words that co-occur with words in green nodes.

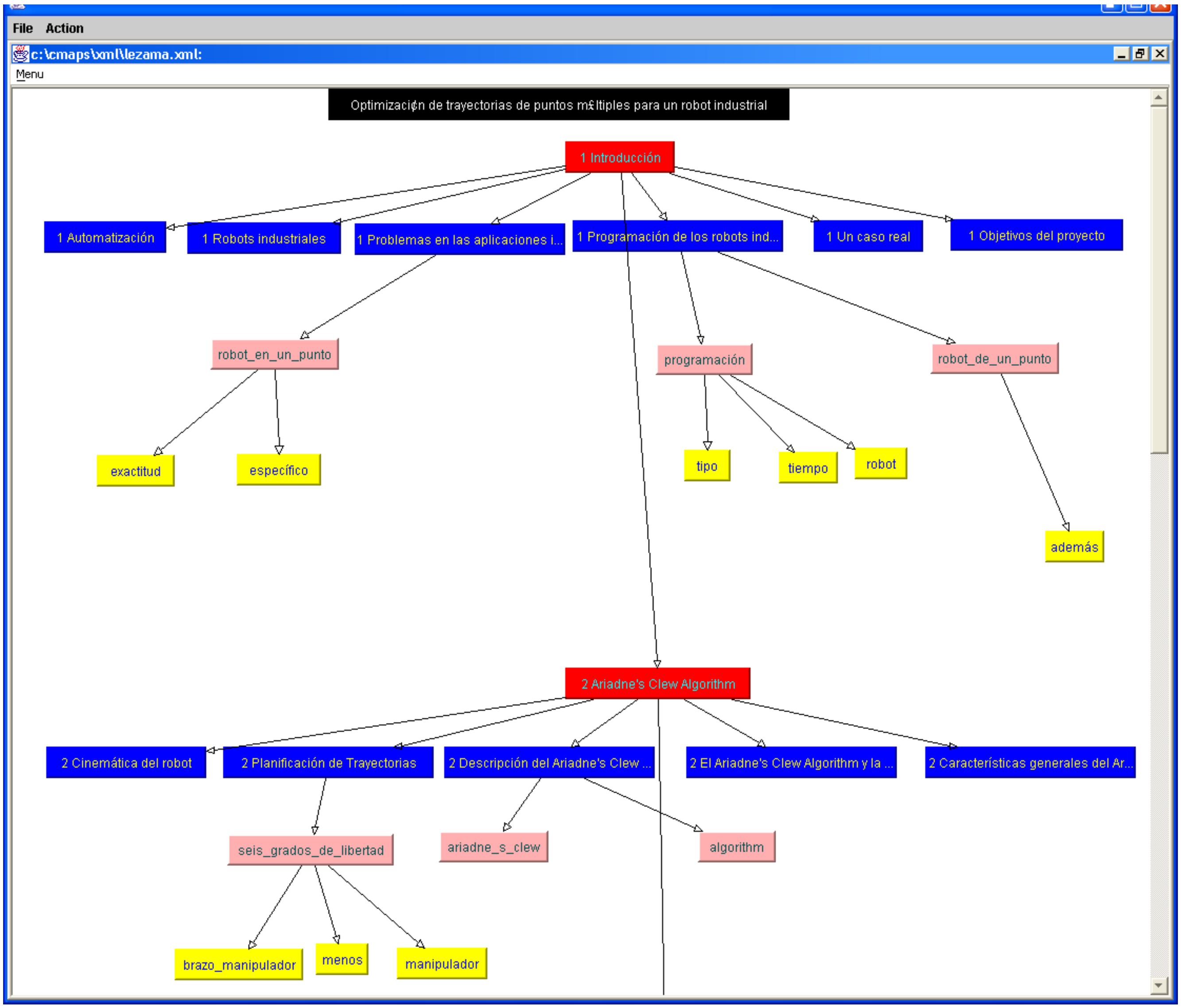


Figure 16: Table of contents map for a Spanish master's thesis. Information obtained from "Optimization of multiple-point trajectories for an industrial robot" (English translation) by Ruth Lezama Morales. Black node is the thesis title, red nodes are chapter titles, blue nodes are section headings, pink nodes are words that co-occur with words in the section headings, and yellow nodes are words that co-occur with words in green nodes.

4.4 Selection of relations

Several methods of finding relations in the text have been evaluated. The first that I have used involves association rules [4]. The following is a formal statement of the problem that association rules were developed to solve [3]: Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called its *TID*. We say that a transaction T contains X, a set of some items in I, if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if c% of the transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has support s in the transaction set D if s% of the transactions in D contain $X \cup Y$.

4.4.1 Fisher's exact test

As an alternative to association rules, Fisher's exact test (left-sided) was evaluated, using the NSP package [11]. The test works like this: assume that the frequency count data associated with a bigram <word1><word2> is stored in a 2 by 2 contingency table as shown in Table 6. Here n_{11} is the number of times <word1><word2> occur together, and n_{12} is the number of times <word1> occurs with some word other than word2, and n_1p is the number of times in total that word1 occurs as the first word in a bigram.

Table 6: 2 x 2 contingency table

	word2	**~word2**	
word1	n_{11}	n_{12}	n_1p
~word1	n_{21}	n_{22}	n_2p
	np_1	np_2	npp

The left-sided test is calculated by adding the probabilities of all the possible two by two contingency tables formed by fixing the marginal totals and changing the value of n_{11} to less than the given value. A left-sided Fisher's Exact Test tells us how likely it is to randomly sample a table where n_{11} is less than observed. In other words, it tells us how likely it is to sample an observation where the two words are less dependent than currently observed. Unfortunately this method assigns the same numerical value to many relations. Counts are used to break the ties. This is very similar to how association rules are found (simply counting occurrences with the most occurrences winning), so in practice it is largely equivalent to association rules.

4.4.2 Dice coefficient

The Dice coefficient is an information theoretic measure similar to mutual information. Assuming that the frequency count data associated with a bigram <word1><word2> is stored in a 2 by 2 contingency table as shown in Table 6, the Dice Coefficient is defined as $\frac{2 \times n_{11}}{np_{1} + n_{1}p}$

4.4.3 Log likelihood

The log-likelihood ratio [32] is a power-divergence test, which puts it in the same category as Pearson’s Chi-Squared test. It measures the deviation between the observed data and what would be expected if <word1> and <word2> were independent. The higher the score, the less evidence there is in favor of concluding that the words are independent. Interestingly, for the dissertations that I looked at, log likelihood gave similar results to mutual information, even though mutual information is an information theoretic measure.

4.4.4 t-score

Given the contingency table in Table 6, the t-score is defined as follows:

$$m_{11} = n_1 p \times \frac{np_1}{npp}$$

$$\text{t-score} = \frac{n_{11} - m_{11}}{\sqrt{n_{11}}}$$

The t-score test gave reasonable results, comparable to association rules. See Table 7. The main advantage to using t-scores is that it is possible to use a "sliding window" approach to finding related concepts, which cannot be done efficiently using current association rule algorithms. The reason for this is that association rule algorithms assume a fixed number of market baskets, not market baskets of variable size. The results in Table 7 are for a window size of 10 (approximately the length of one sentence, to make it comparable to the association rule results). For concept maps made using t-scores, the links are naturally undirected, since if A implies B, then B must necessarily imply A. I produced whole-document maps using t-scores and also using association rules, for comparative purposes. The association rule maps seemed to be better since the most frequent terms had higher degree, which can be used by the map-placement algorithm. Perhaps a different map placement algorithm could be found that would make t-score maps look better, but it is not obvious how that would work. One future study could evaluate t-score maps versus association rule maps, comparing user responses to relations found using association rules and those found using t-scores. Mutual information and log-likelihood also were evaluated, and gave very similar results to t-scores. Table 7 shows the top ten relations found in the master's thesis *Birds of a Feather? How Politics and Culture Affected the Designs of the U.S. Space Shuttle and the Soviet Buran* by Stephen Garber [40]. Note that the association rules include an 'implies' relation (denoted by "<-"). The t-score, mutual information, and Dice coefficient measures merely show bi-directional relation in the text, which means that each term necessarily implies the other (denoted by '<>' in the table).

Table 7: Top ten relations from a master's thesis, using association rules, t-scores, mutual information, and Dice's coefficient. Relations that were selected by more than one measure are in bold.

Association rules	t-score	Mutual Information	Dice coefficient
NASA <- transportation	**energiya<>buran**	**interview<> stephen_garber**	military_characterization<> reluctant_support
NASA <- option	**interview<> stephen_garber**	**range<>capability**	military_characterization<> military
shuttle <- U.S._space	**range<>capability**	**johnson<>soviet_year**	joseph_P<>hopkin
buran <- energiya	NASA<>shuttle	**energiya<>buran**	blueprint<>misleading
soviet <- space_technology	**cross<>range**	**cross<>range**	myriad_american<>blueprint
soviet <- aviation_week	**stephen_garber<> november**	izvestiya<>december	soviet_figures<>chronology
space <- station	**johnson<>soviet_year**	flying<>december	julian_E<>settings
space <- transportation	interview<>november	yaroslav_golvanov<> december	key<>soviet_bibliography
space_shuttle <- decision	**soviet_year<>space**	**stephen_garber<> november**	Soviet_figures<> soviet_bibliography
space_shuttle <-	technology<>culture	**soviet_year<>space**	political_story<>

jenkin			comparative_case

In addition to association rules and t-scores, Dice’s coefficient, Pearson's Chi-Squared, pointwise mutual information, odds ratio, and phi coefficient also were evaluated. These gave surprisingly similar results, which were generally not useful for a concept map. Based on my experiments, I could group the measures into 4 types based on similarity of results.

- group A = association rules, Fisher’s exact test
- group B = t-score
- group C= mutual information, log likelihood
- group D = Dice’s coefficient, Pearson's Chi-Squared, pointwise mutual information, odds ratio, phi coefficient

Groups C and D were not appropriate for a concept map showing the most important concepts, however, since these algorithms mainly select obscure names or phrases throughout the documents, and ignore common ones. One could say these measures emphasize confidence over support, while association rules require high support for rules to be considered. Group D measures could be used to find fixed phrases since they give high weight to words that occur together with high frequency, without regard to support in a document. Currently, though, the application uses only grammatical information to find fixed phrases. Group D also could be used for a researcher looking specifically for uncommon phrases in documents. Experiment 2 used concept maps made with association rules, which I categorize as Group A. I did not conduct a user experiment using concept maps made with a Group B relation, since I determined that using more advanced NLP techniques was a more fruitful research path (see Chapter 5).

4.5 Experiment 2

4.5.1 Motivation of experiment

Since I had developed three design layouts for automatically-generated concept maps, I decided to conduct an experiment to find users' reactions to the three layouts. In October 2004 I ran an experiment to compare the user-perceived effectiveness of whole document maps, chapter level maps, and table of contents (ToC) maps, as described in sections **4.3.1, 4.3.2,** and **4.3.3**. Based on my prior choice of document genre (see Section 1.1), I decided to use ETDs from Virginia Tech as my test collection.

4.5.2 Experimental setup

There were 5 subjects. Each subject skimmed 2 ETDs, randomly selected out of 6 possible ETDs. The median length of the ETDs was 75 pages. All ETDs were in English. These ETDs had been chosen at random from the VT-ETD collection, so each subject was presented with two randomly chosen ETDs.

Users were asked to complete a subject profile detailing their previous use of ETDs. All had a comparable level of experience reading ETDs, so this was not analyzed further. Each user was given 20 minutes to skim each ETD. The users then were asked to complete four tasks for each ETD.

Task 1. Identify the most important concepts in the ETD. Rank those concepts in order of importance.

Task 2. After each user was presented with 2 new sets of concepts (created by the researcher), partitioned into disjoint sets called A and B, each user was next asked to fill in the correct relations (based on the ETD) from concepts in set A to concepts in set B.

Task 3. Each user drew a concept map for the ETD in whatever style they wanted.

Task 4. Each user was shown a concept map automatically generated from the ETD. One was a whole document map, one was a chapter level map, and one was a table of contents map. Each user was asked to rate each of them on a 5-point Likert scale with respect to concept (node) selection, relation (link) selection, and overall usefulness as a summary of the document. Table of contents maps were referred to as "hierarchical" maps so as not to make it obvious how the map was produced, which might have caused bias.

Users then repeated tasks 1-4 for a second ETD.

4.5.3 Experimental results

Task 1

Table 8 has information about the list of concepts the users wrote in Task 1.

Table 8: Task 1 results

Number of	**Average**	**Median**	**Mode**
Concepts/Thesis	10.4	10	10
Words/Concept	6.8	4	3
Characters/Concept	45.2	26	19

Task 3

Table 9 has information about the maps the users drew in Task 3.

Table 9: Task 3 results

Number of	**Average**	**Median**	**Mode**
Nodes	18.1	16	18

Words/Node	2.1	2	2
Characters/Node	16.1	15	8

Task 4

Table 10 lists the average Likert scores given by users in Task 4. The score were from 1 (lowest) to 5 (highest).

Table 10: Summary of results for task 4

Concept Map Type	Concept Selection	Link Selection	User-perceived usefulness
Whole Document	2.5	2.6	1.9
Chapter Level	2.1	2.5	1.9
Table of contents	3.6*	3.3*	3.6*

* = significantly better than both other types in paired t-test, p < .05

Table 11 lists the significant results of the paired t-tests comparing table of contents (ToC) maps, chapter level maps, and whole document maps.

Table 11: Comparison results of experiment 2

	Better than	at which characteristic	p-value
ToC maps	whole document maps	concept selection	0.017
ToC maps	chapter maps	concept selection	0.000
ToC maps	whole document maps	link selection	0.045
ToC maps	chapter maps	link selection	0.037
ToC maps	whole document maps	overall usefulness	0.002
ToC maps	chapter maps	overall usefulness	0.001

None of the relationships between whole document maps and chapter maps, involving concept selection, link selection, or overall usefulness, approached significance.

4.5.4 Discussion of results

- Users greatly preferred the ToC concept maps to whole document or chapter maps. This could be because they have a more organized structure, which was accentuated by the color-coding.
- Some users also commented that the colors in the ToC maps were helpful to them. This agrees with previous CMap research from the TCU group [25]. However, it

happens that users did not find color-coding based on part-of-speech (as used in the chapter and whole document maps) to be very helpful.

- The most elaborate map had 33 nodes, averaged 17 characters per node, and used 10 of the 11 possible colors.
- Many maps included cycles and therefore were not trees. This supports the conclusions of [101], where it is suggested that cycles are needed to represent some patterns in the real world, i.e., if concept maps are to model the real world properly, cycles should be allowed. The whole document and chapter-by-chapter maps will sometimes have cycles in them, while the Table of Contents maps will not.
- There seemed to be a disconnect between subjects' idea of what a concept is and what a node should contain. The average length of a concept in Task 1 was 6.8 words or 45.2 characters, while the average node in Task 3 had only 2.1 words or 16.1 characters. In fact the longest concept in Task 1 had 31 words, while the longest node in Task 3 had only 8 words. One explanation is that the GetSmart concept map editing tool [71], by default, only shows the first 11 characters of a node text, and to see beyond this the user must resize the node. Interestingly, the mode of the node text was 8 characters, which means the text would fit entirely within the displayed part of the node. Perhaps this encouraged students to write shorter node text than they otherwise would have. However, the average length (16.1) is still longer than the 11-character limit.
- Four of the 10 maps used color as a way to divide the map into different sections. The other 6 maps did not use color.

- The whole document and chapter-by-chapter maps needed major improvement if user criticisms were to be avoided.

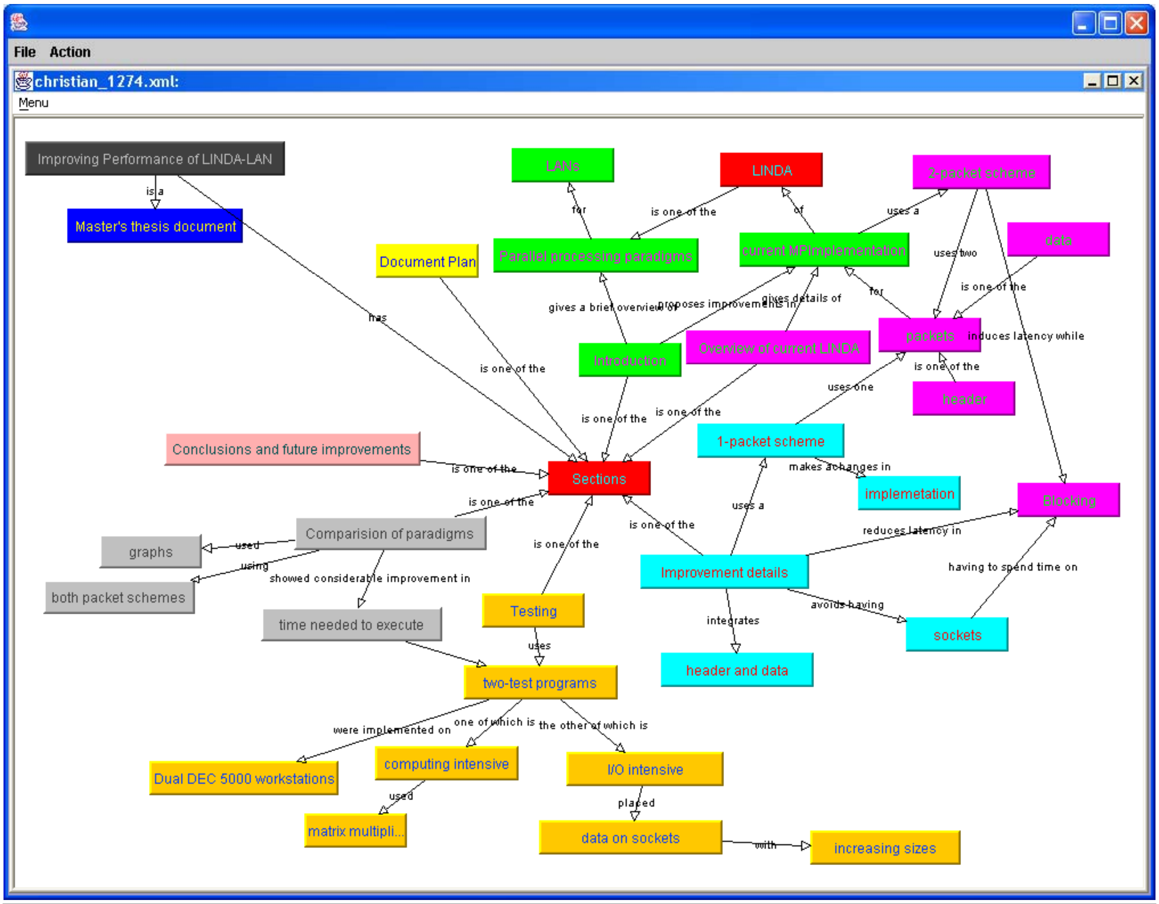


Figure 17: Concept Map drawn by subject. Information obtained from "Improving Performance of LINDA-LAN" by Jason L. Christian.

5 Using NLP-aware tools for concept map creation

Since I have experimental results suggesting that concept maps can be used in addition to abstracts as a cross-language summarization tool, and since I have results suggesting the preferred type of map for certain tasks, our main goal is to automatically generate concept maps from English documents, and automatically translate the nodes and links into Spanish. However, the techniques that I used in previous experiments, based on using only minimal NLP processing, were not very successful. Thus I incorporated state-of-the-art NLP, combined with an ontology, to produce much more informative maps. As my test collection I chose computer science ETDs at Virginia Tech, since there are a number of these (about 200) readily available without access restrictions.

In the next sections I summarize how I have produced concept maps of English documents using NLP and a computing ontology, using a combination of tools.

5.1 Use of Relex to automatically generate maps of English ETDs

Relex was developed at VettaLabs in Brazil, under the direction of Ben Goertzel at Virginia Tech. It is based on ideas from Minipar by Dekang Lin at the University of Alberta [64]. Relex is a system that translates syntactic dependencies into a graph of semantic primitives [120], by means of template matching algorithms. Relex also detects implied quantities, normalizes passive and active forms into the same representation, and assigns tense and number to sentence parts. Relex must be used in conjunction with a dependency grammar parser. The current Relex version uses Carnegie Mellon

University's link parser [106] to extract the base dependencies. Most of Relex's morphological analysis is performed using WordNet's morphological functions [34]. The link parser often breaks when presented with unknown, multi-word entities, such as "Prime Minister Gerhard Schroder" or "Cisco Systems Inc.". Instead of adapting the link parser to better handle such phenomena, the authors of Relex have taken an alternate approach, in which they tag named entities and replace them with simple identifiers. For instance, instead of feeding the sentence "Donald Macloud Jr. is going to Belo Horizonte to assume the post of dean at Abraxas University", the system feeds "ID1 is going to ID2 to assume the post of dean of ID3" into the parser. All substituted entities IDs are treated as noun phrases. After the sentence is parsed and converted into a semantic primitive graph, the IDs are restored back to full entity names, and each entity is assigned its proper tag (in the example above, Person, Location, and Organization, respectively). Relex comes with a set of tools to extract and draw paths between named entities in the semantic primitive graph. Figure 18 shows the link parser output, semantic-primitive based relationships, and inter-entity paths resulting from Relex analysis of the sentence "Prime Minister Renate Schimidt searched England for Cisco Systems".

```
Prime Minister Renate Schimidt searched England for Cisco Systems.

Entities
Person = Prime Minister Renate Schimidt
SocialOrganization = Cisco Systems
GeographicalLocation = England

Link Parser Linkage
+---------------------------Xp--------------------------+
|                    +--------MVp-------+               |
+----Wd----+----Ss----+----Os----+       +---Jp---+      |
|          |          |          |       |        |      |
LEFT-WALL genericID1 searched.v genericID2 for.p genericID3 .

Semantic Primitives
for(searched,Cisco Systems)
_obj(searched,England)
_subj(searched,Prime Minister Renate Schimidt)

Graph
```

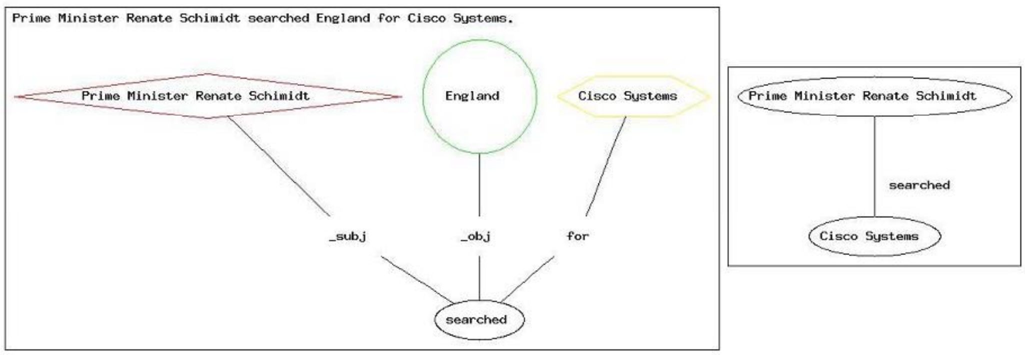


Figure 18. Various stages of the Relex Processing Pipeline (drawn by Ben Goertzel. Reproduced by permission. See Appendix VII.)

Relex generates entity maps of documents. Each node is labeled with a concept (represented by one or more words), and each link is labeled with one or more linking words, which usually are verbs but also can be nouns used as modifiers. The fact that links can sometimes have nouns as labels instead of only verbs or prepositions is the main reason that the authors of Relex refer to the maps as entity maps instead of concept maps.

Relex is implemented within IBM's UIMA framework [50]. This framework describes a series of design patterns, interfaces, and metadata to implement, combine, and deploy analysis capabilities. The default entity tagger used with Relex is based on the Another Nearly-New Information Extraction (ANNIE) system, distributed as part of the

General Architecture for Text Engineering [30]. This tagger was customized to better deal with multi-word and non-English names.

The base version of Relex has two customized entity taggers, one specialized for the biomedical domain and another for world news and finance. Neither of these is appropriate for our purposes regarding computing ETDs. In order for Relex to be able to recognize computing terms as entities, I needed a comprehensive (and consistent) source of terminology. I found this in The Ontology Project [23], in development at Villanova University and sponsored by ACM. This project has divided computing into 21 topic level domains, and provides a hierarchy of terms that go from one to six levels deep for each of these main topic areas. The ANNIE extractor needs to recognize when an instance of a class in the ontology appears in the text. I accomplished this by selecting various nodes in the ontology and encoding "gazetteers" – lists of individual terms that would be the surface representations of that node in the document. Since the computing ontology has about 900 leaf nodes, it would be very time-consuming to write gazetteers for all of them. Therefore I selected a few areas of computer science for which Virginia Tech has a large number of ETDs (i.e., digital libraries, human-computer interaction, virtual environments) and wrote gazetteers for only these (200+) nodes.

Since Relex currently only works for English documents, this forces us to use Spanish-speakers as the target audience. Although it would be useful for Relex to help generate maps of Spanish documents (or more likely Portuguese ones, since it was developed in Brazil), there are no plans to do this at present. Part of the reason is that the English link parser is an integral part of the application, so replacing it would require considerable effort.

Currently I am using an English->Spanish wordlist [88], compiled at the University of Maryland, to translate nodes that consist of a single word. Obviously this could be improved upon. Research done at IHMC involves translating single word nodes by combining concept map link information with WordNet to determine which synonym set (synset) a particular node should be associated with [19]. Once this is determined, one can cross-reference that synset with its nearest equivalent in the Spanish EuroWordNet to find the best translation of that word. Incorporating this type of system into my concept map translation system is left as future work.

There are other, simpler ways to make the Spanish concept map more appealing. One useful technique is to translate the map from English to Spanish and then to compute the ideal node placement, instead of using the node placement from the English map. I do this by translating the concept map while it is still in Graphviz's Dot format (developed at AT&T) [39], which contains no node-placement information. Translating the map after the node placement information has already been computed can lead to problems, such as the case when the Spanish translation is much longer than the original English word. An example is the noun 'reuse', which might translate to Spanish as 'reutilización'. The larger node in the Spanish map might lead to the 'reutilización' node, obscuring another nearby node. Thus the appearance of the English and Spanish maps will differ slightly since Graphviz's node placement algorithm is not directly under the user's control.

5.2 Summarizing large concept maps

In some cases, the map generated by Relex will be larger than what is optimal for a user to read. Our experiment 2 indicated that people prefer maps with no more than 20 nodes.

The question is, how do I find the most important concepts in the maps, so that I can leave these and remove the less important ones?

I first apply ideas from concept map theory, regarding what makes a good concept map. Novak and Gowin put great emphasis on the careful selection of link labels between concepts. Ideally each concept-link-concept triple should form a meaningful proposition ([87], p. 34). Also, ideally a concept map is made with respect to a specific question called a focus question; thus it should have no unconnected sections (p. 3). Interestingly, Relex does not always meet the first criterion. In some cases Relex identifies that two nodes are related, but cannot find an appropriate word to label the relation between them. Since these relations are not meaningful propositions, nor very helpful to readers, I remove these unlabeled relations to prune the graph. Thus, if a node is only connected to other nodes by unlabeled relations, then that node will be removed from the graph.

To implement the second criteria, I use a simple heuristic. If a map has more than 20 nodes, I count the degree of each node, and remove all nodes with degree=1. Since these nodes detract from the overall structure of the map, removing them makes the map look more like a traditional concept map.

In the case of very large maps (hundreds of nodes), I need additional techniques to reduce the size of the graph. In the case that the graph still contains more than about 20 nodes, I can apply more advanced techniques. One example is to compute the TF-IDF of each node in the graph and only keep the nodes with a TF-IDF above a certain threshold. For this I compute the term frequency (TF) based on just that dissertation, and the inverse document frequency (IDF) based on the entire VT-ETD computer science collection (about 200 documents). Another possibility is to compute the weighted mutual information [89] for each relation, and only keep the most highly weighted relations.

5.2.1 Example of English automatic concept maps

Below I present some automatically generated concept maps of chapters of dissertations in English. I present maps for 3 chapters from Hussein Suleman's dissertation, since it is an example of a dissertation with terms that are well-covered in my computer science named-entity extractor. Then I present some chapters from other VT-ETDs in computing. All of these maps have been processed by a simple "summarizer" heuristic, which removes all relations which are not labeled, and also removes all nodes with degree less than 2. This heuristic is applied only once to each map, so in some cases, due to nodes being deleted, there will be remaining nodes with degree less than 2. The maps are taken from screenshots of IHMC's CmapTools software [18].

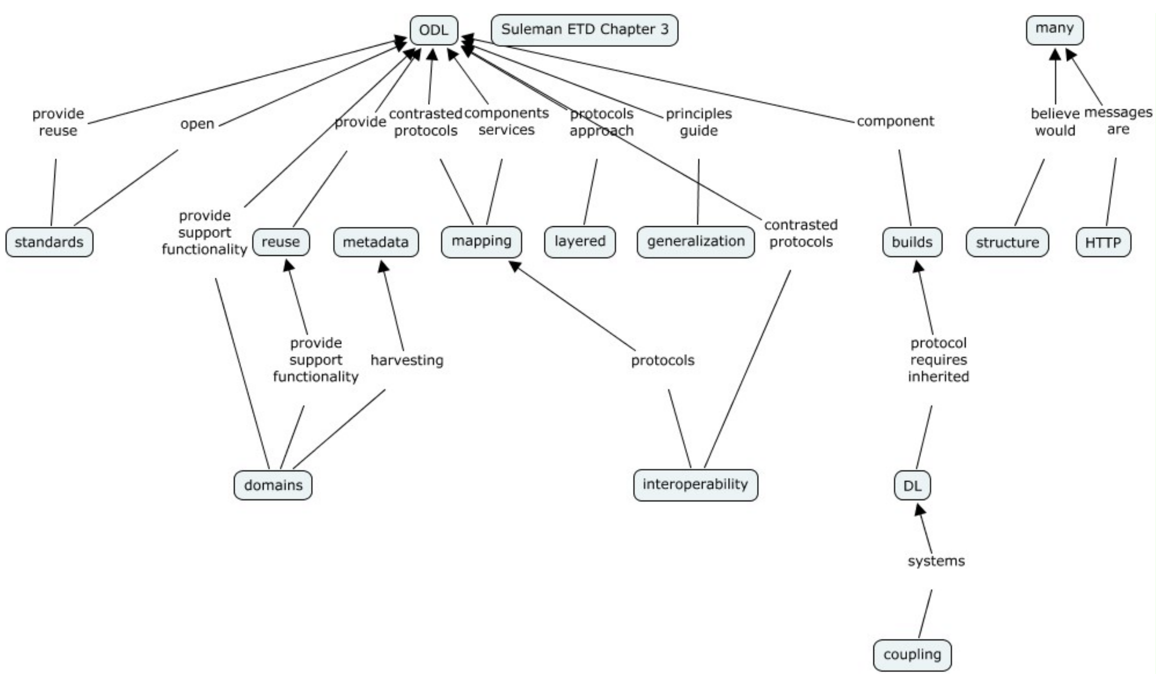


Figure 19: Automatically drawn concept map. Information obtained from 'Open Digital Libraries' by Hussein Suleman, Chapter 3

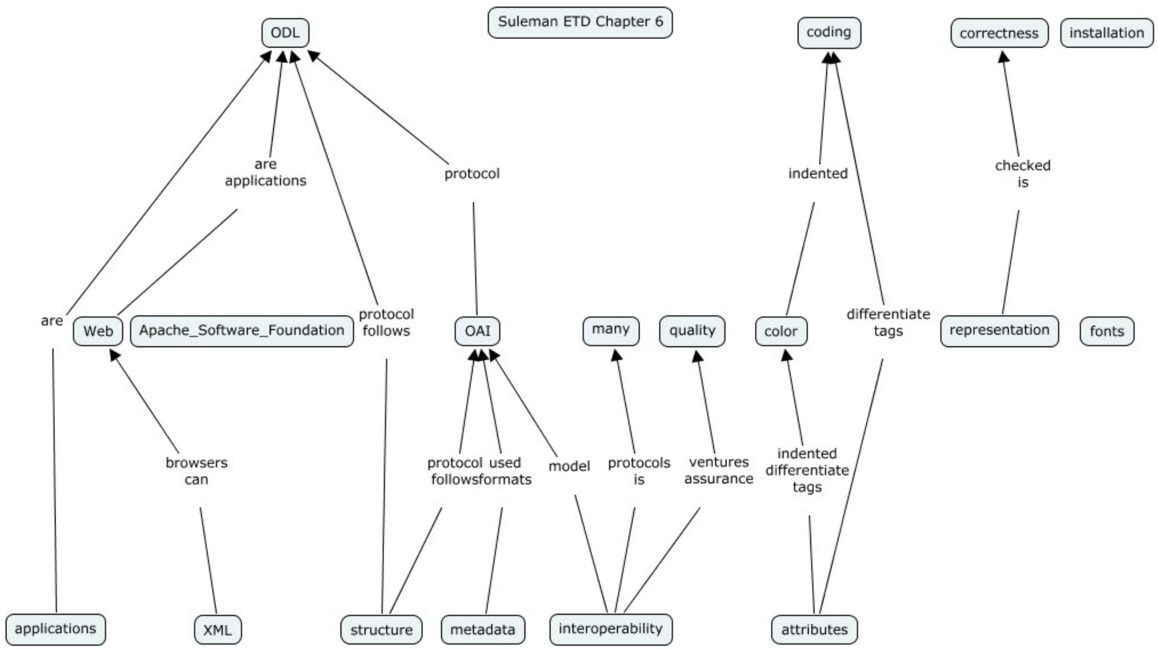


Figure 20: Automatically drawn concept map. Information obtained from 'Open Digital Libraries' by H. Suleman, Chapter 6

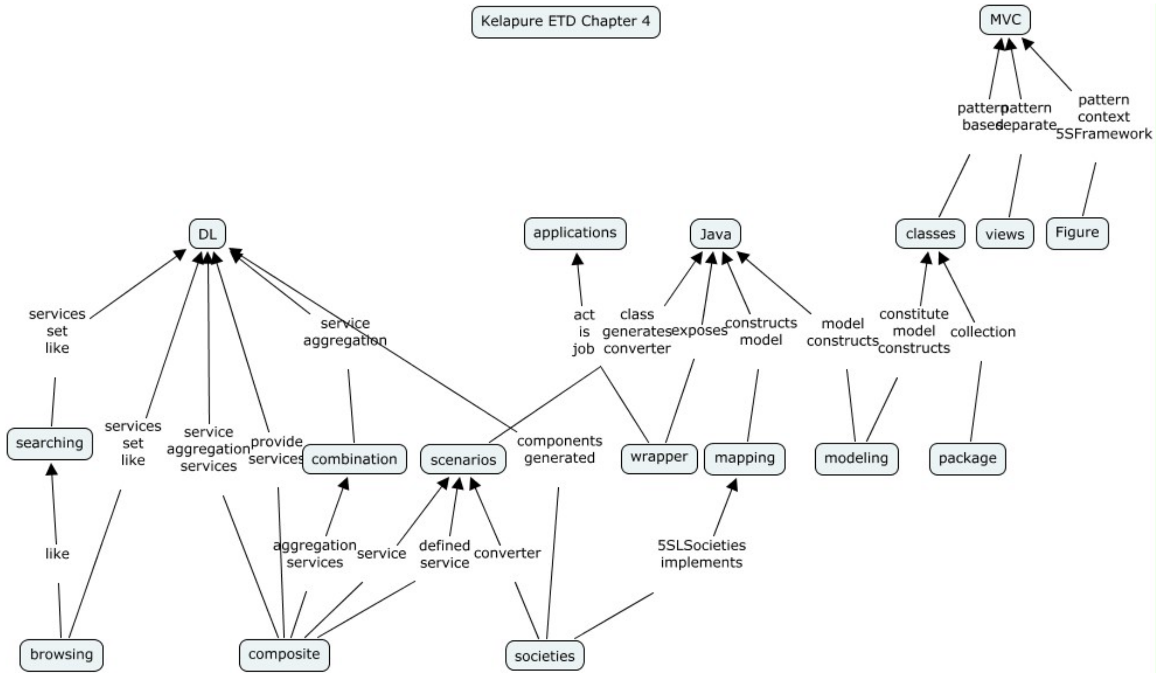


Figure 21: Automatically drawn concept map. Information obtained from "Scenario-Based Generation of Digital Library Services" by Rohit Kelapure, Chapter 4

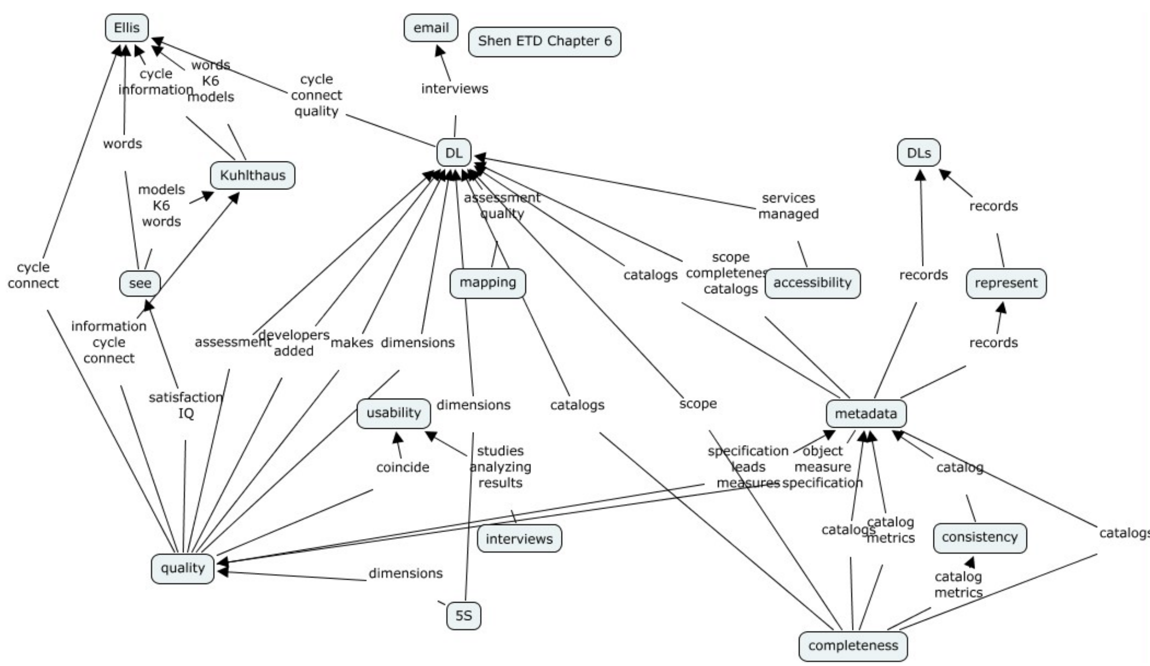


Figure 22: Automatically drawn concept map. Information obtained from "Applying the 5S Framework To Integrating Digital Libraries" by Rao Shen, Chapter 6

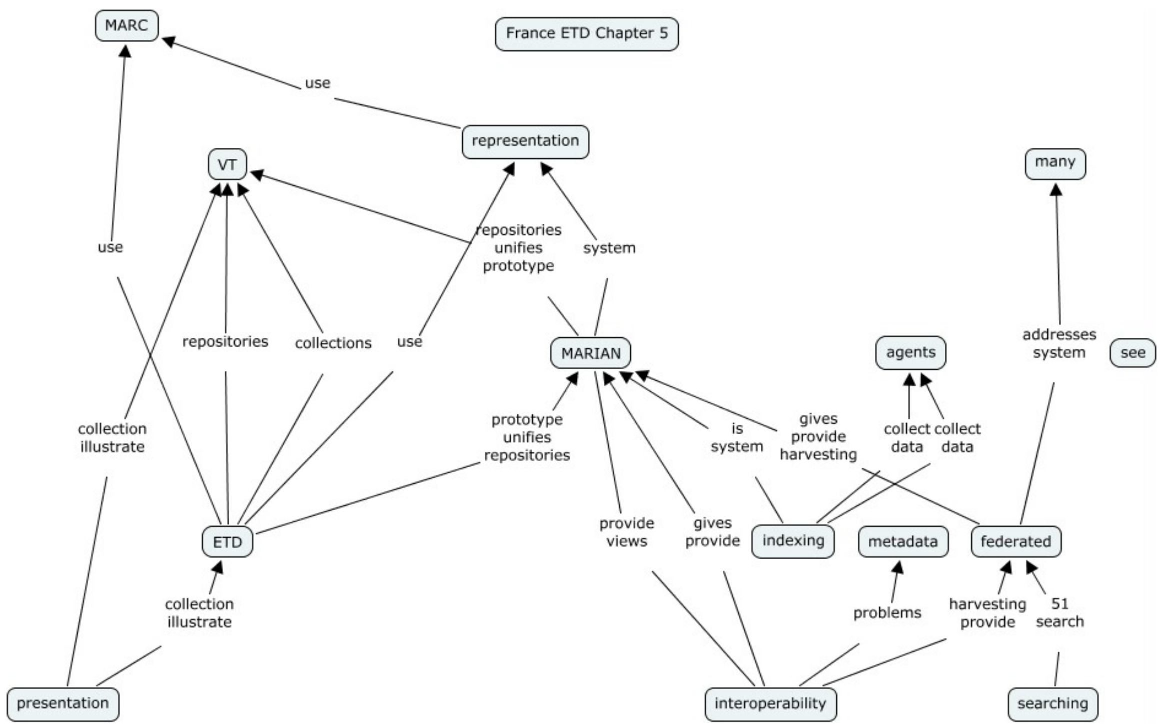


Figure 23: Automatically drawn concept map. Information obtained from 'Effective, Efficient Retrieval in a Network of Digital Information Objects' by Robert K. France, Chapter 5

5.2.2 Using TF-IDF to summarize maps

Below is a concept map automatically generated from Chapter 2 of Hussein Suleman's dissertation. It contains 65 concepts and 109 relations, and is too convoluted to be of much use to a human reader.

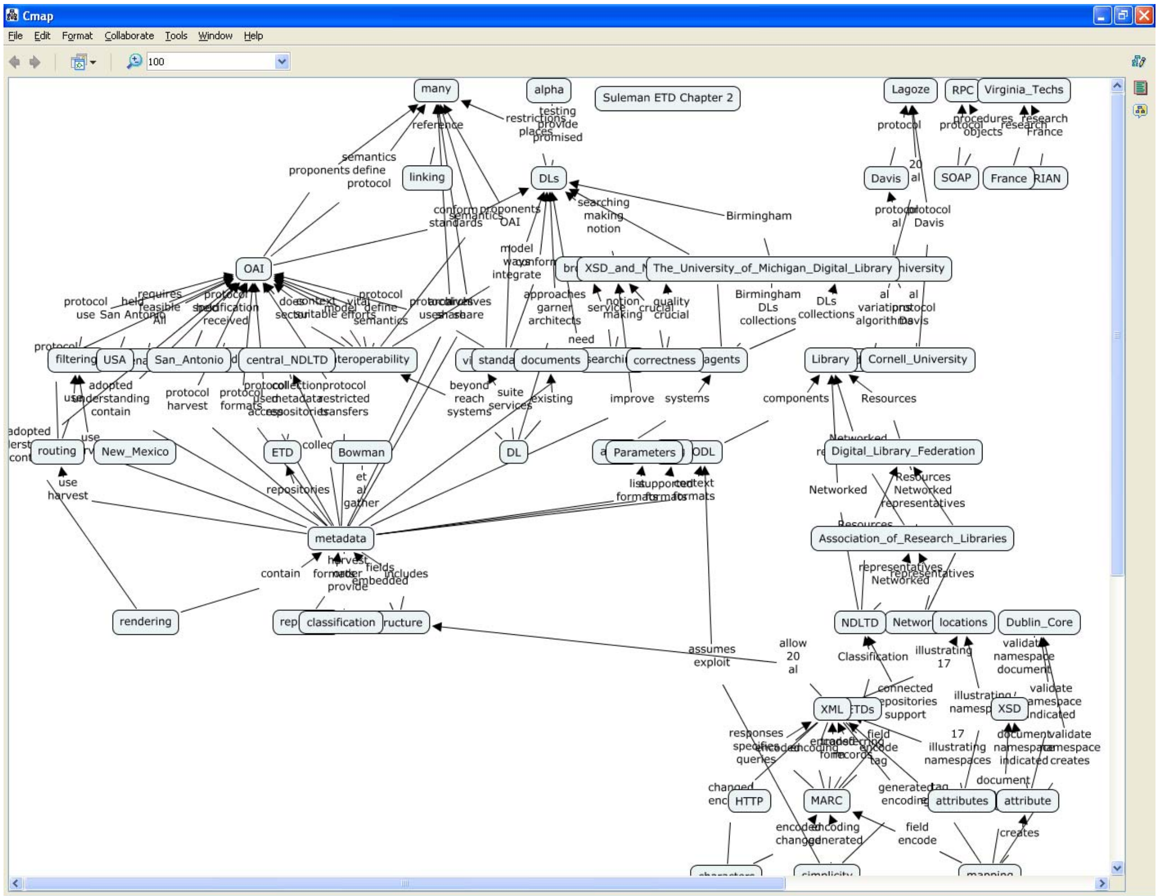


Figure 24: Automatically drawn concept map. Information obtained from "Open Digital Libraries" by H. Suleman. The concept map has 65 concepts.

Below is a map that has been pruned using TF-IDF values. To prune the graph, I selected the 20 highest TF-IDF values, and then only used relations in which both nodes were in this top 20 list. This produced a map with 30 relations. I also tried pruning the map by selecting all relations that had at least one node in the top 20 TF-IDF ranking, but this only reduced the size of the map slightly, from 109 relations to 89.

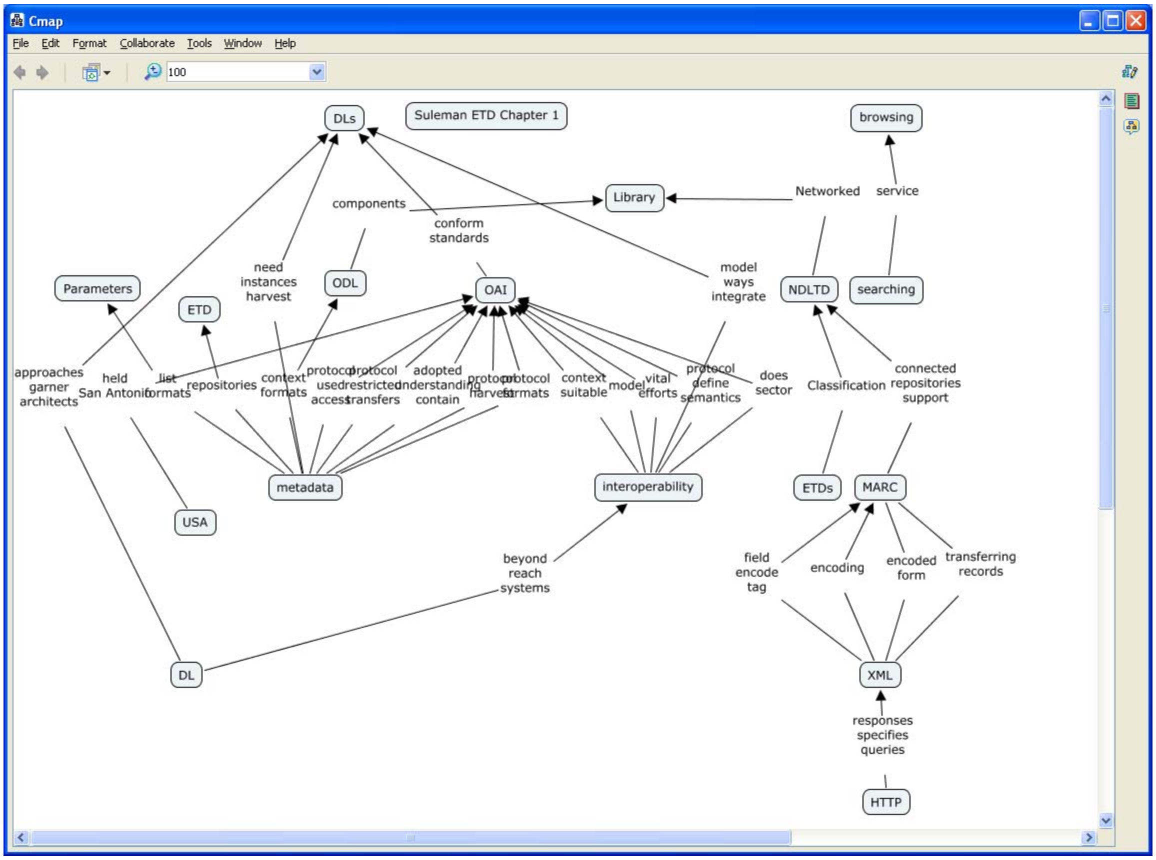


Figure 25: TF-IDF pruned version of previous concept map. Information obtained from "Open Digital Libraries" by H. Suleman.

5.3 Combining Relex with Table of Contents maps

One problem with the concept maps automatically generated using Relex is the lack of an overall structure. In Experiment 2, I found that users liked maps that used the table of contents as a framework more than they liked less-structured maps. Thus I can carry out a similar process with the Relex maps. That is, I can use the chapter and section headings as a skeleton for the map, and find important concepts in each section that relate to the section heading. Figure 26 shows a part of a concept map drawn in this manner. This particular graph is quite large, since the ETD has 9 chapters, and each chapter has between 4 and 14 sections. So it is not possible to show the entire map in one screenshot. Users are required to use the CMapTools interface to browse the map.

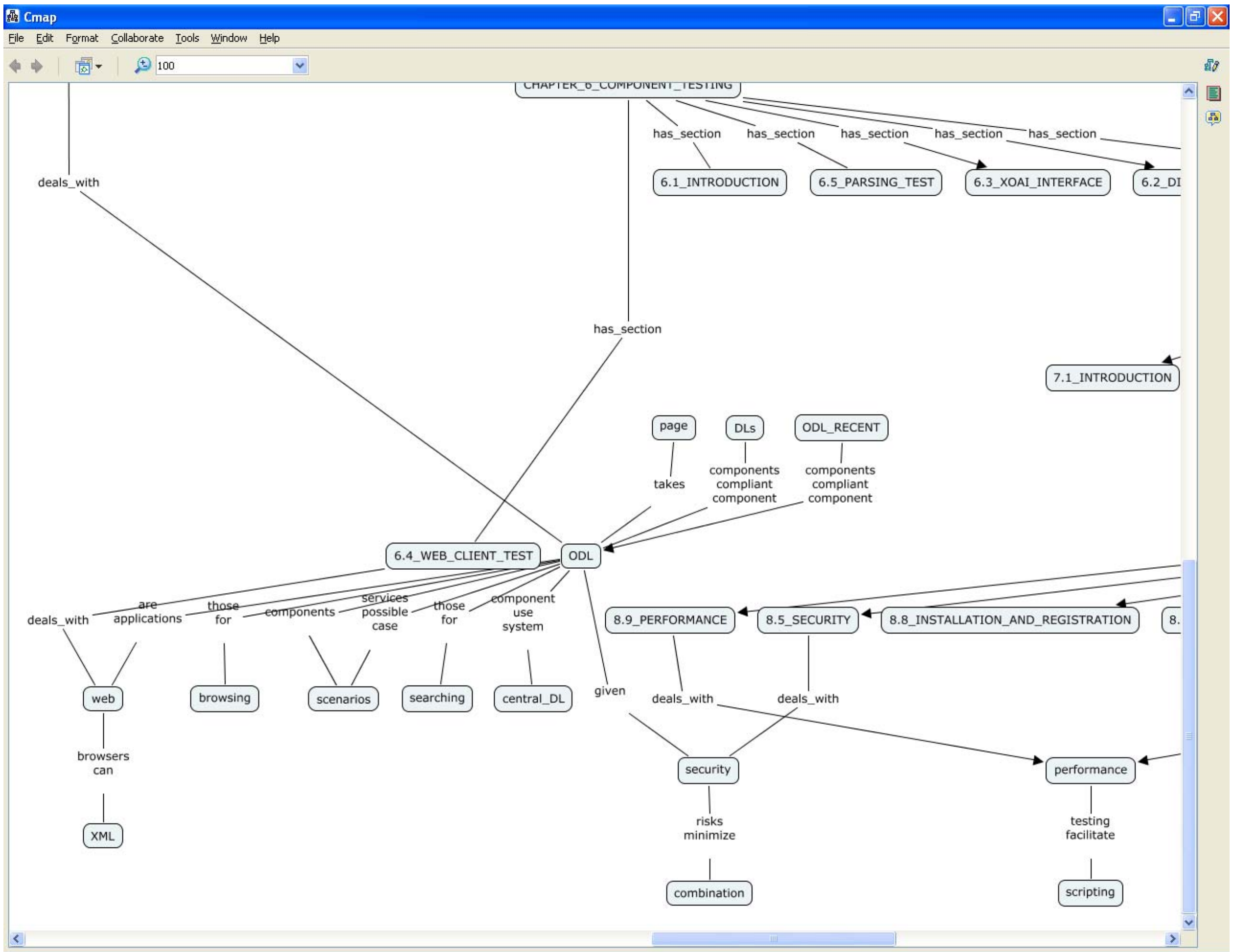


Figure 26: Section of a Table of Contents graph with additional entities found by Relex. Information obtained from "Open Digital Libraries" by H. Suleman.

The concept map below was generated from a dissertation by O. Sornil. It is clear from the map that index partitioning schemes are an important topic in the work.

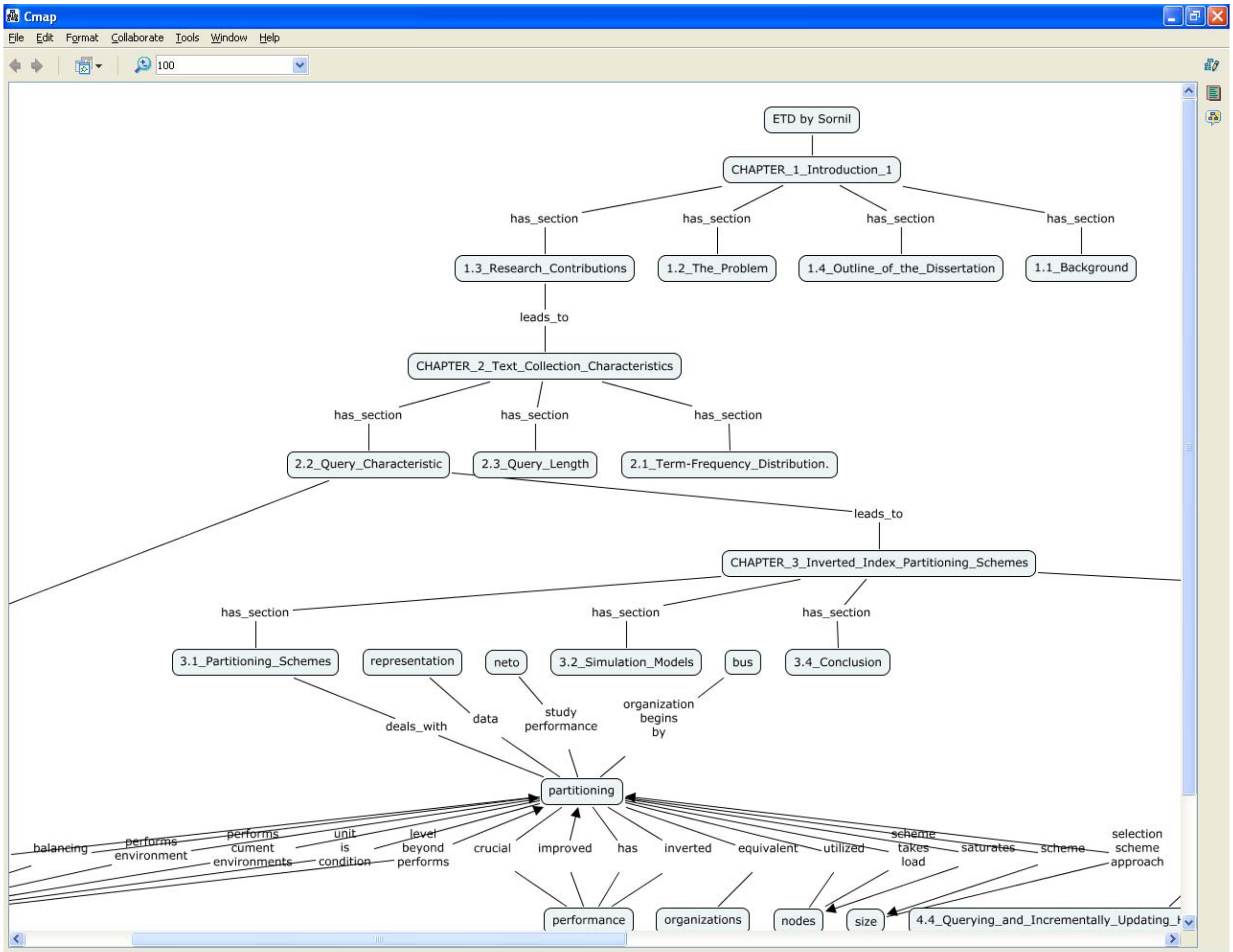


Figure 27: Section of a Table of Contents graph with additional entities found by Relex. Information obtained from "Parallel Inverted Indices for Large-Scale, Dynamic Digital Libraries " by Ohm Sornil.

5.4 Experiment 3

Before continuing with cross-language experiments, I decided to test the improvements that I had made to English-only concept maps. The simplest way to do this was to essentially re-run my experiment 2, utilizing two promising concept map styles that I had developed. The first style was the Relex-generated chapter maps, combined one after the other to show the entire dissertation. The second style was the Table of Contents+Relex merged maps. The study called for participants to evaluate automatically generated concept maps based on several subjective criteria, including node selection and link selection.. The concept maps were displayed in CMapTools and were formatted using CMapTools' new AutoLayout feature. There were 10 subjects.

Participants completed a pre-testing survey about their degree of familiarity with computer science and concept maps. Participants were given an electronic thesis (ETD) in the computer science discipline from the VT-ETD collection (http://scholar.lib.vt.edu/theses/) and asked to read or skim it for 15 minutes. All theses used were those with unrestricted access. After skimming the ETD, they were shown the two different styles of automatically generated concept maps based on the ETD. Half of the participants saw the Relex-only generated maps first, and half saw the Table of Contents+Relex generated map first. This was to account for learning effects of seeing the concept maps.

The participants then were asked to rate various characteristics of the concept maps on a five-point Likert scale. They answered subjective questions about the node selection, link selection, usefulness of references back to original ETD (available by hovering over link text), and overall satisfaction with the concept maps. They also had space to supply additional comments about the concept maps. The participants then repeated the process on another computer science ETD. Thus each participant saw two sets of concept maps, four in total. I compared the Likert responses for both types of concept maps to determine which style users preferred.

The results of the experiment are summarized below. Since there were 10 participants, and each saw 2 ETDs, there are 20 observations for each style of map.

Table 12: Results of Experiment 3, Relex-only Maps

Style A (Relex-only Maps)

	Node Selection	Link Selection	Relationships important	Hovertext helpful	Overall
	3	2	3	3	2
	5	4	4	4	4
	2	2	2	2	3
	3	2	1	1	1
	4	4	4	5	4
	5	4	4	5	5
	4	2	3	2	3
	2	3	3	2	2
	3	4	3	3	1
	3	4	3	3	1
	2	3	2	2	2
	4	3	3	3	3
	4	4	4	4	4
	5	3	3	5	4
	3	3	3	3	2
	4	3	2	3	3
	2	3	2	3	3
	3	4	4	3	4
	3	2	2	1	1
	2	2	2	2	2
Average:	3.3	3.05	2.85	2.95	2.7

Table 13: Results of Experiment 3, Table of contents + Relex Maps

	Style B (Table of contents + Relex Maps)				
	Node Selection	Link Selection	Relationships important	Hovertext helpful	Overall
	4	4	3	4	**4**
	3	2	3	3	**3**
	4	4	4	4	**4**
	3	3	4	5	**4**
	5	3	4	5	**5**
	4	3	3	4	**4**
	3	4	3	4	**5**
	4	4	4	2	**4**
	4	4	4	4	**5**
	4	4	4	3	**5**
	3	3	2	4	**3**
	3	4	3	4	**4**
	5	5	5	5	**5**
	5	5	3	5	**5**
	5	4	4	3	**5**
	5	5	4	3	**5**
	4	4	3	3	**4**
	4	4	4	2	**4**
	3	3	2	1	**2**
	3	3	2	2	**2**
Average:	3.9	3.75	3.4	3.5	4.1
p-value:	**0.015**	**0.006**	**0.012**	**0.030**	**0.000**

As can be seen from

Table 13, users preferred the Table of Contents + Relex style maps (style B) in all five categories (Node Selection, Link Selection, Importance of Relations, Hovertext, and Overall Usefulness for Summarizing). The differences were statistically significant ($p < 0.05$) for all five categories.) More information about Experiment 3 can be found in Appendix III.

5.5 Experiment 4

Experiment 4 was largely a repeat of Experiment 3, with multi-layer concept maps and more participants. There were three main motivations in running Experiment 4.

1. Based on the user judgments on the Table of Contents maps, it was unclear if adding the Relex relations added any value over just drawing the table of contents.
2. Laying out the entire map of an ETD as one concept map was overwhelming for users, according to their comments. Users were required to scroll down to see past the first two chapters, which was distracting. Therefore I decided to make a 'nested' map, in which the top level had nodes only for the title and chapter titles. Concept maps for each chapter are attached as resources to the nodes labeled by the chapter titles. Thus students can concentrate on each chapter at a time, without having to take in the entire ETD at once.
3. I needed to know which type of concept map I should use for further experiments. The options were maps which simply drew out the table of contents, maps consisting entirely of terms found by Relex, or a combination of the two.

The three layout types of concept maps tested were:

A) Maps generated only using the Table of Contents (ToC) maps

B) Maps with only Relex-found terms from the computing ontology (plus chapter titles)

C) Maps which combine types A and B above.

These three layouts are illustrated below. All three have the same overview view.

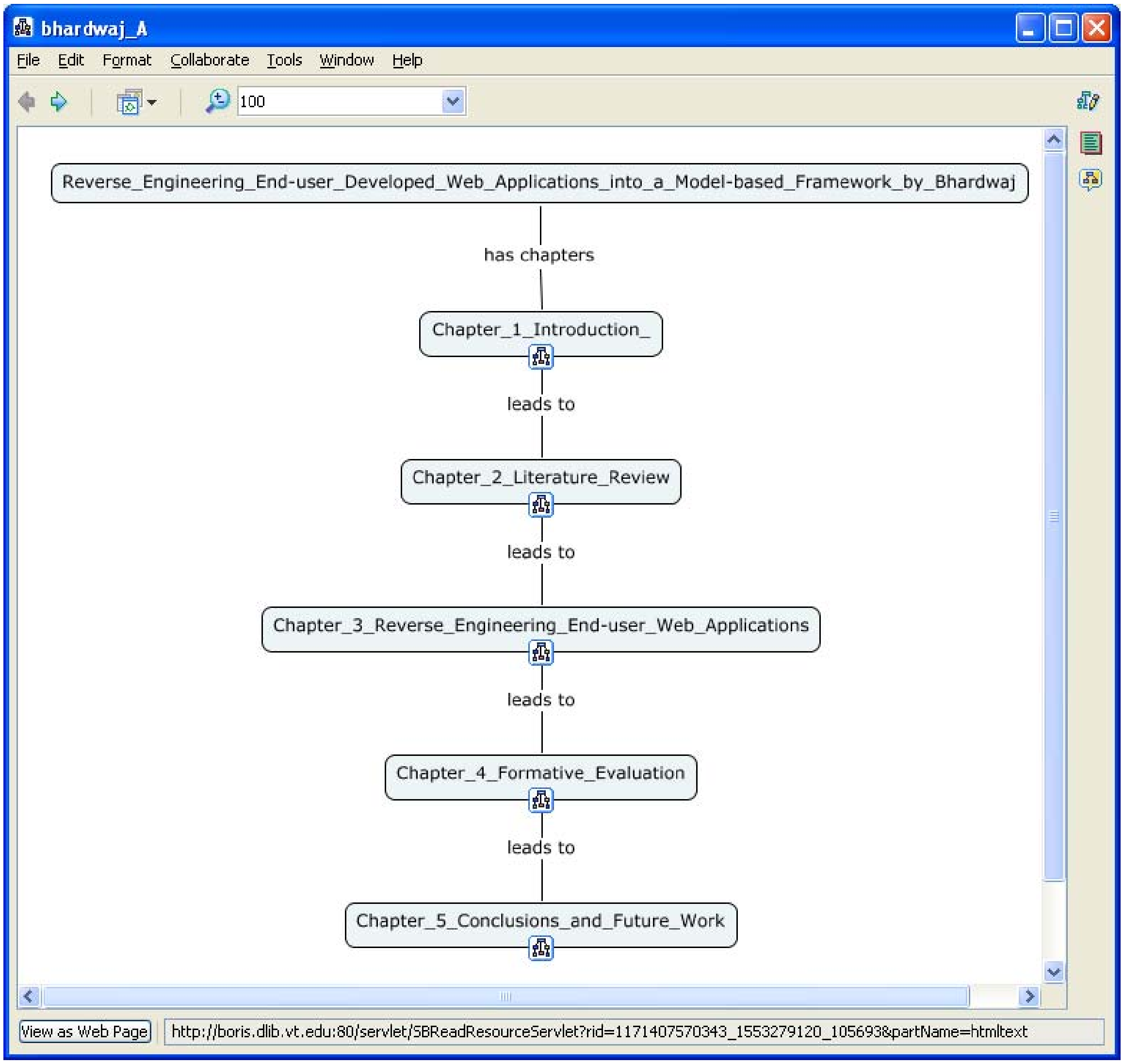


Figure 28: Overview map for an ETD. Information obtained from "Reverse Engineering End-user Developed Web Applications into a Model-based Framework" by Yogita Bhardwaj.

By clicking on the square icons at the bottom of each node, the user can open the concept map for each chapter. The ToC-only view is shown in Figure 29.

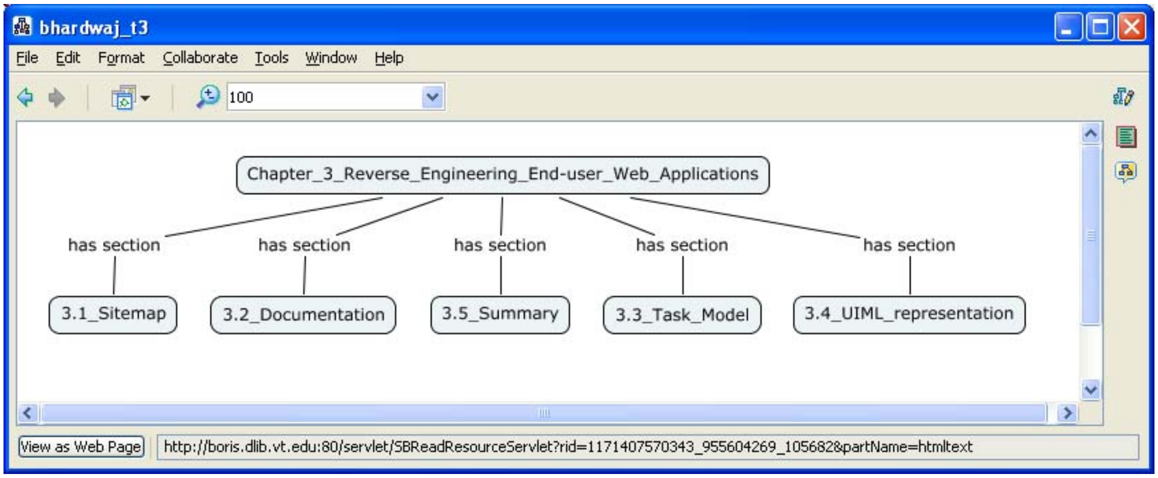


Figure 29: Table of Contents map of an ETD. Information obtained from "Reverse Engineering End-user Developed Web Applications into a Model-based Framework" by Y. Bhardwaj, chapter 3.

The layout with terms found by Relex is shown in Figure 30.

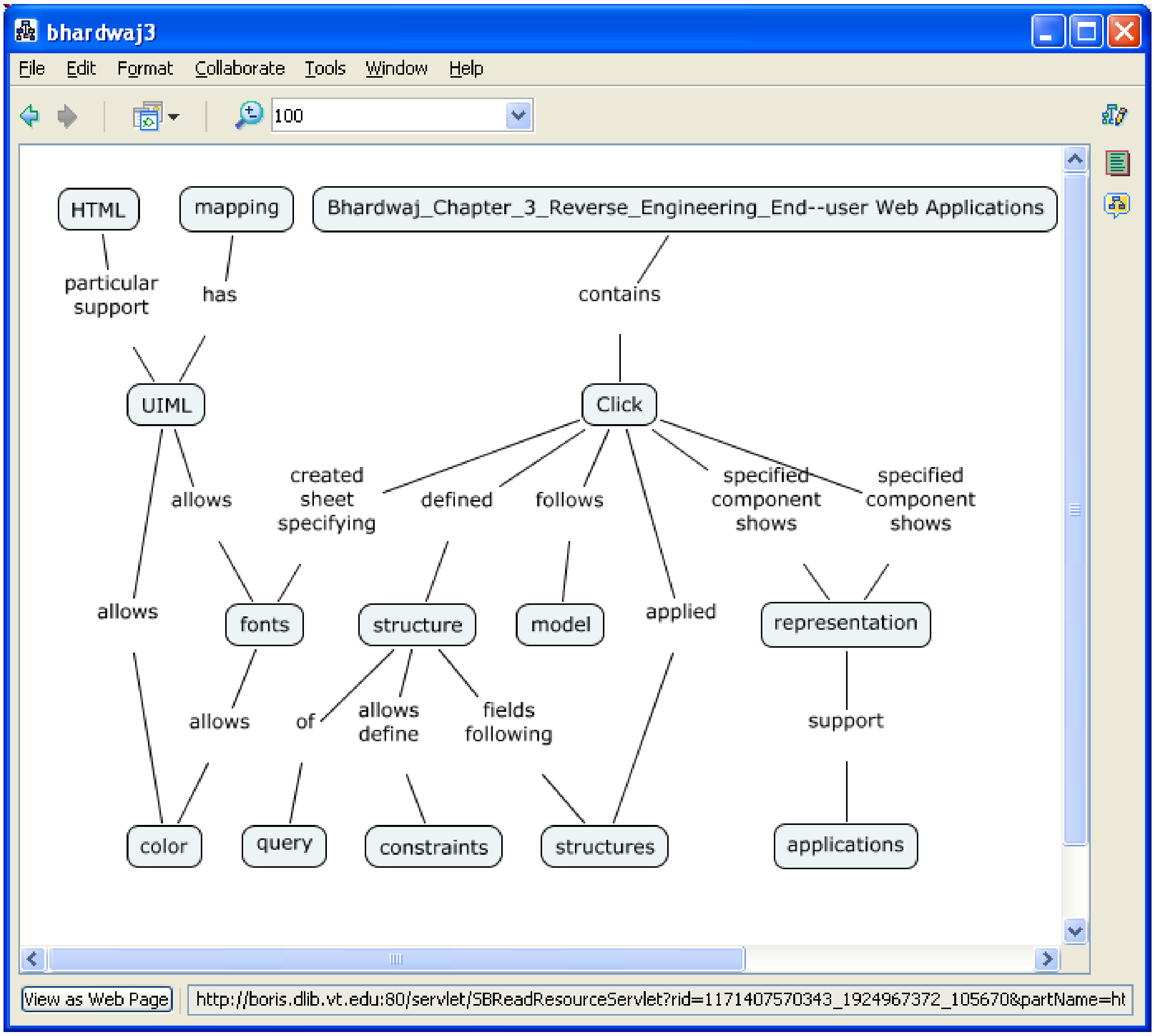


Figure 30: Relex-term map of an ETD. Information obtained from "Reverse Engineering End-user Developed Web Applications into a Model-based Framework" by Y. Bhardwaj, chapter 3.

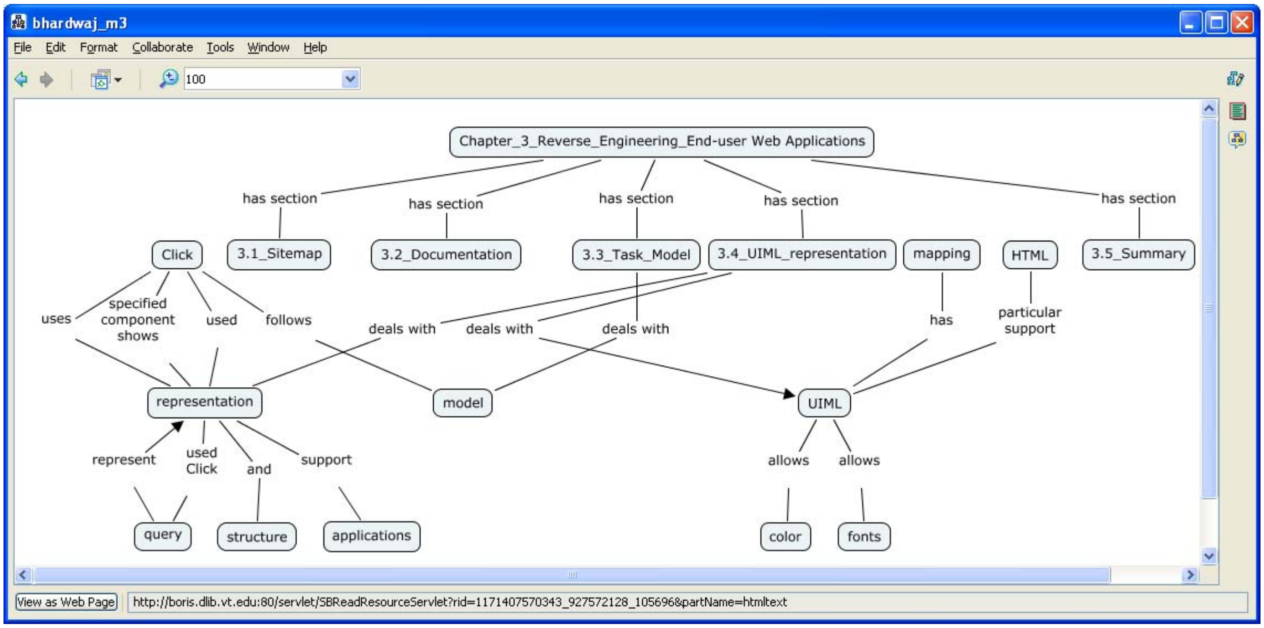


Figure 31: Table of Contents + Relex terms concept map. Information obtained from "Reverse Engineering End-user Developed Web Applications into a Model-based Framework " by Y. Bhardwaj, chapter 3.

The questions and experimental design were essentially the same as for experiment 3. Thirty-five subjects participated in the experiment. Subjects were asked to rate the concept maps on a 5-point Likert scale based on node selection, link selection, whether the relationships were important in the ETDs, helpfulness of hovertext, and overall usefulness in determining what the ETD is about. I performed paired t-tests comparing the layouts A, B, and C against each other based on these criteria. Users were presented with only 2 of the 3 concept map layouts in the interest of allowing them to finish in one class period. Presentation order was randomized, making 6 possible orders. These were AB, BA, AC, CA, BC, and CB. Users were randomly assigned to one of these six presentation orders. The results are summarized in the table below.

Table 14: Comparison of Table of Contents (ToC) and Relex-only Maps

Layout	Node selection	Link Selection	Relationship Selection	Hovertext	Overall
ToC-only	**2.29**	2.81	**2.00**	**2.86**	2.81
Relex-only	**3.19**	3.24	**2.86**	**3.71**	3.52

Differences in bold were significant at the p=0.05 level.

Table 15: Comparison of Table of Contents and Table of Contents + Relex maps

Layout	Node selection	Link Selection	Relationship Selection	Hovertext	Overall
ToC-only	**2.83**	3.71	**2.63**	2.92	**2.87**
ToC + Relex	**3.50**	3.42	**3.29**	2.96	**3.67**

Table 16: Comparison of Relex-only and Table of Contents+Relex maps

Layout	Node selection	Link Selection	Relationship Selection	Hovertext	Overall
Relex only	3.46	2.92	3.35	**3.54**	3.46
ToC + Relex	3.20	3.58	3.23	**3.08**	3.42

As one can see, both Relex-only and ToC+Relex concept maps were rated higher by users than the ToC-only concept maps. Relex-only and ToC+Relex maps were rated almost the same by users across the board, with the only significant difference being that users thought that the hovertext for the Relex-only maps was more informative. For ToC+Relex maps, there are two types of hovertext. For a link between a chapter and a section, or between a section and a Relex-found terms, the hovertext shows the first 200 characters of the text of that section. For relations between Relex-found terms, the hovertext shows the sentence in the original thesis where both of these terms occur. Why users seemed to prefer only the hovertext between Relex-terms is unclear from user comments.

For Experiment 5, I needed to choose one concept map style to present to users. Either Relex-only or ToC+Relex would have been valid choices, since both were rated higher than the ToC-only layout in 3 of 5 categories. I chose the ToC+Relex layout, because it was the only layout for which users gave significantly higher ratings on the ‘Overall’ question than the ToC-only maps.

6 Using ETDs as a resource to automatically discover phrase translations

To make the concept maps useful as a cross-language tool, they must be automatically translated from English to Spanish (since Relex only works for English documents). While widely-available machine-readable dictionaries might be sufficient to translate one-word concepts in the maps, they will not contain the technical phrases found in most ETDs. Thus some way to find translations of technical phrases must be found. I used two methods to do this:

1. To jumpstart the process, I extracted the top two levels of the ACM CC2001 hierarchy (550+ phrases), and then had a computer science expert, Fernando das Neves, translate them into Spanish. This was completed in January 2006.
2. To find more phrase translations, I used English and Spanish ETDs in the computer science domain as a comparable corpus to find these phrase translations.

6.1 ETD collection development

For our corpus, I needed a collection of Spanish ETDs relating to our target domain, which is computer science. With assistance from Scirus (www.scirus.com), I acquired 2580 Spanish ETDs which they had harvested from NDLTD. These ETDs did not explicitly list which field they were in, nor even the department in the university that they came from, so I selected 417 of them which contained the text “computacion” and/or “informatica” (ignoring accented characters) for our computing-related collection. The ETDs from Scirus contain the first 20 KB of text, formatted normally, but the fulltext has all punctuation removed. Thus I could not use sentence boundaries for part-of-speech tagging on the fulltext. To use the unformatted fulltext, I implemented a custom n-gram based part-of-speech tagger to find noun phrases in the collection. I conducted initial tests to determine if using the phrases found in the fulltext was worthwhile, or if I could ignore the fulltext and only use the properly-formatted summaries. I determined that, even given the imperfect part-of-speech tagging (and hence noun phrase extraction) when using the fulltext, I still found Spanish phrases that I would not discover otherwise. Thus I used both the fulltext and the summaries for our collection.

I also acquired 78 ETDs from the computer science department of Universidad Autónoma de Mexico (UNAM) in Mexico City, and 89 ETDs from the Universidad de las Americas (Puebla, Mexico) which were in computing or closely related fields. Since I needed a wordlist, I acquired an 47,000 entry bilingual dictionary from the University of Maryland that had been mined from parallel corpora. I enhanced this dictionary with content from four computing-related websites that have Spanish/English glossaries. These were www.telefonica.es (a telecommunications company, listing 232 phrases), www.css.qmul.ac.uk (Queen Mary University of London, listing 82 phrases), www.jilt33.com/dti (developed by BOYC Technology, listing 840 phrases), and http://www.geocities.com/Athens/Forum/2323/glosario.html (a collection of 473 computing and telecommunication phrases). Since there was overlap between the phrases provided by each of the websites, these resources actually provided me with 1,451 English to Spanish phrase translations.

I extracted all noun phrases with 2 to 5 words from the union collection. This resulted in 730,869 Spanish phrases. I extracted 2-5 word English noun phrases from Virginia Tech's CS-ETD collection (195 documents). As another test collection of English phrases, I used 1971 phrases from The Ontology Project [23], which contains computing-related vocabulary from the ACM Computing Classification Scheme [1], the Australian

Computer Society [107], the Curriculum Recommendations of the ACM, IEEE-CS, AIS [2], and the German Accreditation for Informatics Programs [7].

6.2 Phrase translation algorithm

Lopez-Osteñero et al., in [68], describe a phrase-translation algorithm suitable for use with comparable corpora. Since the algorithm does not use deep language knowledge, it could work with a variety of language pairs. I have implemented an English-to-Spanish version of Lopez-Osteñero's phrase translation algorithm to find the phrase translations that I need to translate the concept maps. Since Lopez-Osteñero's algorithm requires a bilingual dictionary, I used the aforementioned wordlist from University of Maryland. The inputs and outputs of the algorithm are shown in Figure 32.

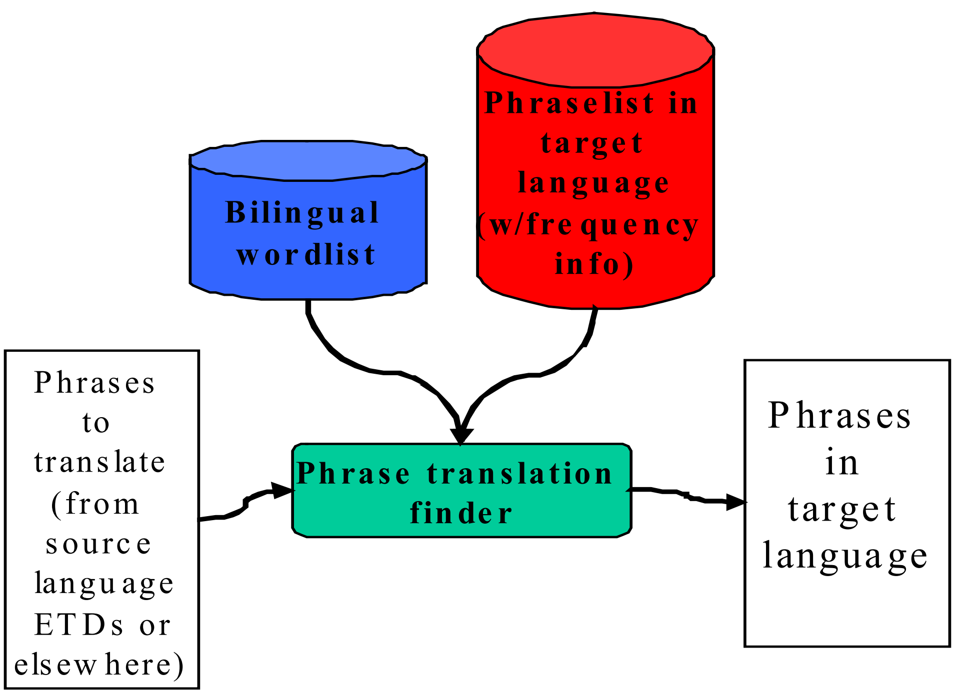


Figure 32: Flow diagram of Lopez-Osteñero phrase translation algorithm.

The algorithm works as follows.

Given an English phrase P, a Spanish phrase list S_{list}, and a bilingual dictionary D,

- For each word W in P
  - Find in D the set of Spanish words that can be translations for W, call this W_{trans}
  - For each word *s* in W_{trans}
    - Find all phrases in S_{list} containing *s*, put into set $PossTrans_w$
- Perform intersection of all $PossTrans_w$, producing *RTrans*
- Pick the member of *RTrans* that has the highest frequency in a suitable target language corpus.

I made a number of changes to the algorithm. One simple change is that the original paper only discussed 2 and 3 word phrases. I extended the algorithm to handle 4 and 5 word phrases.

I found that sorting the words in the phrase based on number of translations into the target language (fewest translations first) greatly decreases the size of the intersection. This improves runtime by a significant though constant factor but does not change the asymptotic complexity of the algorithm. On a test run looking for translations of 1162 English phrases, word re-ordering ran in 5 hours instead of 7.2 hours, an improvement of 1.4x. Though I only ran the English-to-Spanish case, the Lopez-Osteñero algorithm can be used in both directions.

This optimization works on the assumption that an English word with 2 translations into Spanish will appear in fewer Spanish phrases (in our phraselist) than an English word with 4 translations into Spanish. While this might be approximately true, there are certainly exceptions. For instance, according to our bilingual wordlist the English word “of” can be translated as 5 Spanish words, “de”, “a”, “hacia”, “por”, and “con”. It is possible that for a given phrase another English word might have more than 5 possible translations. However, since all of these Spanish words appear in so many phrases (“de” alone appears in 210,734 phrases), the word “of” should always be placed last when trying to find Spanish phrases that are translations.

Continuing the example of the word “of”, the table below shows how many phrases the possible translations of “of” appear in.

Table 17: Number of phrases containing possible translations of the word “of”

Spanish Word	Number of phrases with this word
de	210,734
a	36,407
hacia	302
por	10,297
con	13,544
Total (simple addition)	271,284

Since one phrase could contain 2 or more of these words, the total of 271,284 phrases is not exact and is merely an upper bound. However, it is easy to compute and serves as a fair approximation of the total number of phrases. Thus I modified the Lopez-Ostenero algorithm a second time to include an index that, for a given Spanish word, keeps track of how many phrases it appears in. I use this index to sort the words in each

English phrase, so that the first word can be in the least possible Spanish phrases and the last word can be in the most possible Spanish phrases. This allows me to ignore many Spanish phrases which in fact have no possibility of being a translation of the English phrase. On a test run using the same 1162 phrases that were used in the previous run, runtime was 1 hour 46 minutes. This is a 4x speedup of the original Lopez-Osteñero algorithm.

6.3 Testing the algorithm

Using this algorithm, my code looked for translations of 1971 English phrases from The Ontology Project in the list of 730,869 Spanish phrases. It found translations for 631 of these (32%). This low percentage is not surprising since The Ontology Project contains phrases from all areas of computing, whereas our Spanish collection is from a very small number of universities which do not conduct research in all of these areas. The correctness of the translations was assessed by two translators. The first expert is a native Spanish speaker in Argentina with a PhD in computer science. The second expert is a native English speaker who is a high school Spanish teacher. The results are summarized in the table below.

Table 18: Expert judgments on phrase translations into Spanish.

Expert	Correct	Partially Correct	Incorrect
Expert 1	490 (77.7%)	51	90
Expert 2	583 (92.4%)	40	8

'Partially correct' phrases had minor errors that probably could be corrected by post-processing. Some examples are 1) the English phrase is singular, but the Spanish phrase is plural, or 2) the Spanish phrase has nouns and adjectives that do not agree with respect to being singular/plural. The difference in the evaluations between the two Spanish experts shows the importance of using translators who are familiar with the subject matter, in addition to being fluent in Spanish.

In connection with the following experiment (experiment 5), I used this phrase translation algorithm to translate English phrases that appeared in the automatically-generated concept maps. The list of 155 English phrases, and the Spanish translations that were found in the phraselist, can be found in Appendix VI.

7 Cross-language experiment with Universidad de las Américas

To test my key hypotheses about concept maps, I conducted a cross-language experiment with the Universidad de las Americas (UDLA) in Puebla, Mexico. Experiment 5 was a study involving automatically-generated concept maps and human-written abstracts, conducted with students in a class taught by Dr. Alfredo Sanchez.

For the experiment, my three hypotheses were as follows:

1) Automatically-generated concept maps can be a useful summary of an ETD.
2) Automatically-generated concept maps can augment abstracts in helping subjects determine if a document is relevant to an information need.
3) Automatically-generated concept maps can be translated via MT "well enough" that they can be used as a cross-language tool.

7.1.1 Design of experiment

Twenty-two subjects from UDLA participated. There were six treatment conditions, with all subjects seeing all six conditions, making it a fully within-subjects design. Subjects were asked six questions, presented in varying order, with one question in each treatment condition. The experimental materials were based on 30 dissertations from the Virginia Tech CS-ETD collection, all originally in English. For 15 of these, the abstracts and concept maps were translated by a professional translator familiar with IT/computing. For the other 15, the abstracts and concept maps were translated by machine translation (MT). For the machine translation condition, the abstracts were translated by Systran. The original dissertations, and concept maps, and some sample questions are available (in English) at: http://purl.oclc.org/NET/VTDLIB/CMAP/GENERATION/EXAMPLES.

The six questions which were used in the experiment are listed in Table 19.

Table 19: Questions used in Experiment 5

Question	English Translation
¿Cuál o cuáles tesis tratan de clasificación de texto?	Which thesis(es) deal with text classification?
¿Cuál o cuáles tesis mencionan 'Pensar en voz alta' como un tipo de evaluación?	Which thesis(es) mention 'Think-aloud' as a type of evaluation?
¿Cuál o cuáles tesis hablan acerca de la teoría 5S de bibliotecas digitales?	Which thesis(es) talk about the 5S theory of digital libraries?
¿Cuál o cuáles tesis tienen el concepto de escalabilidad?	Which thesis(es) have the concept of scalability?
¿Cuál o cuáles tesis tienen información sobre "cadena de Markov"?	Which thesis(es) have information about 'Markov chains'?
¿Cuál o cuáles tesis tienen el concepto del espacio de trabajo?	Which thesis(es) have the concept of 'workspace'?

For this experiment, I decided not to use the CMapTools representations of the concept maps, and instead used a simpler implementation using AT&T's Graphviz software. This produced JPEG images, embedded in HTML pages. There were two reasons for this:

1. According to Dr. Sanchez, the subjects at UDLA had never used CMapTools before, so they would require a training period before they could use the tool. Since I could not be physically present to conduct a training and answer the

questions they would have, I decided that using a more familiar JPEG/HTML interface would avoid errors caused by lack of familiarity with the interface.

2. Due to a design decision by IHMC, the writers of CMapTools, concept maps stored in IHMC's CXL format (a type of XML) are not treated as concept maps by the application. Instead they are treated as XML files, which are displayed like text files. In order for a CXL file to be viewed as a concept map, the user must "import" the file using the CMapTools interface. Thus, for an ETD with an overview map and 6 chapter maps, the user (in this case, the researcher since this must be done in advance) has to import 7 files to turn it into a concept map viewable by the application. Since this experiment involved a fairly large number of ETDs (30), it would be too time consuming to produce concept maps this way. IHMC is aware of this issue and they indicate there will be a fix in the next version of CMapTools.

Another difference between the concept maps used in this experiment and in Experiment 4 was that these did not include the hovertext feature. Although it is possible to make clickable regions of a JPEG image which point to the sentence from the original document (in fact I implemented such a feature as a test), I opted not to do this since the original document is in English, and I did not have sufficient resources to have all of these sentences translated into Spanish. The layout of Experiment 5 required that everything shown to the subject be in Spanish, so that subjects who can read English fluently did not have an advantage.

The following figures are of concept maps that were used in Experiment 5. The English versions are shown for comparison purposes. All materials presented to the subjects in Experiment 5 were in Spanish.

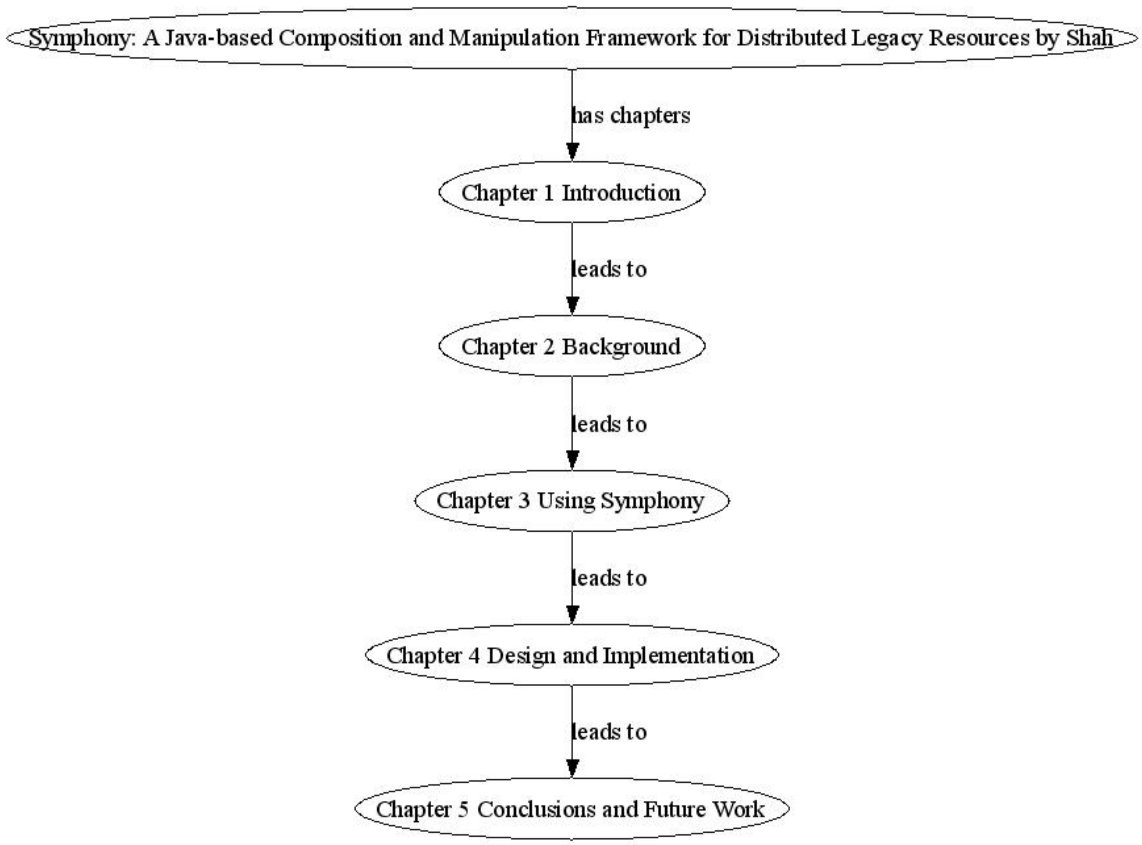


Figure 33: Overview concept map of an ETD. Information obtained from "Symphony: A Java-based Composition and Manipulation Framework for Distributed Legacy Resources" by Ashish B. Shah.

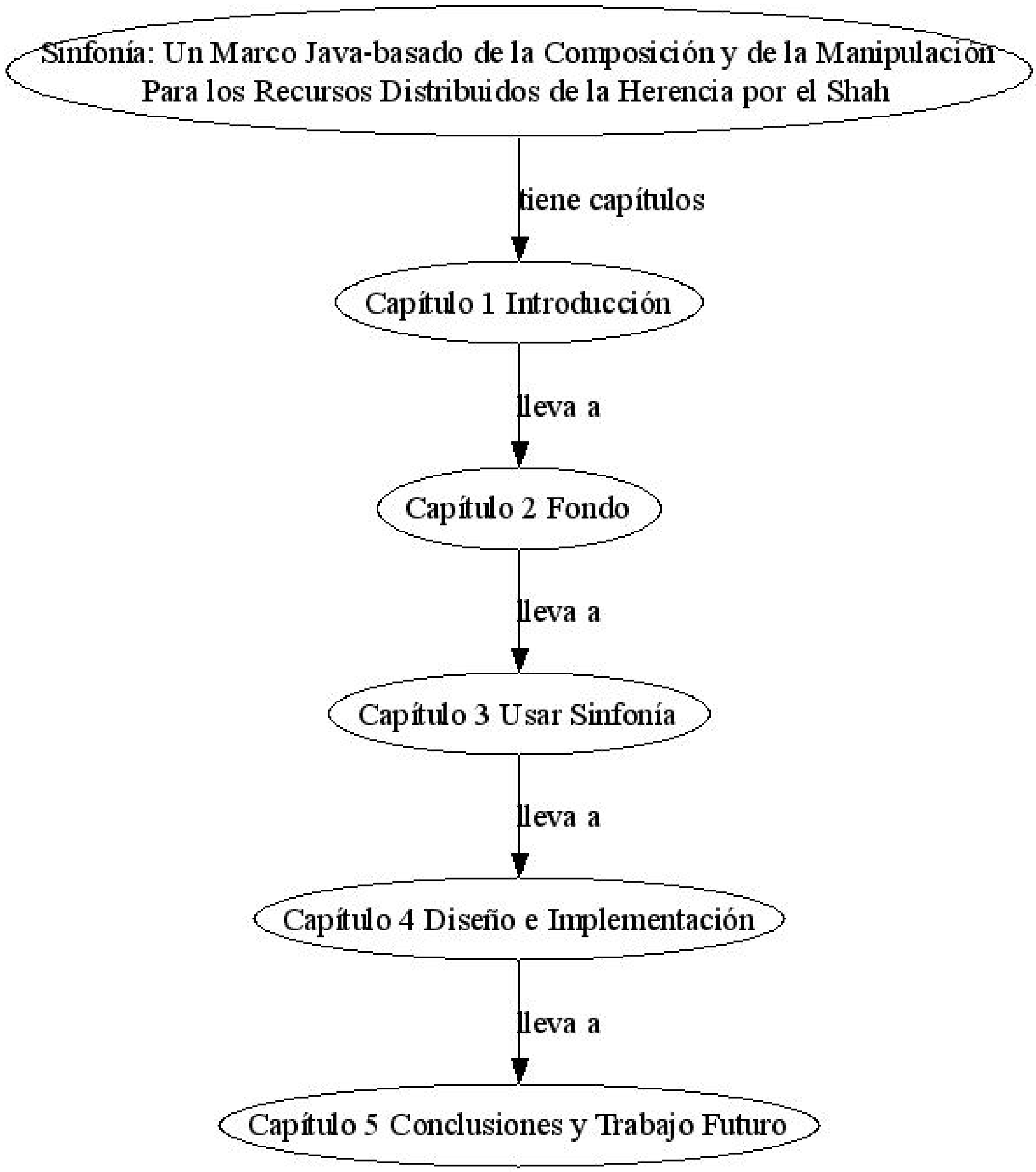


Figure 34: Overview concept map of an ETD, translated automatically into Spanish. Information obtained from "Symphony: A Java-based Composition and Manipulation Framework for Distributed Legacy Resources" by A.B. Shah.

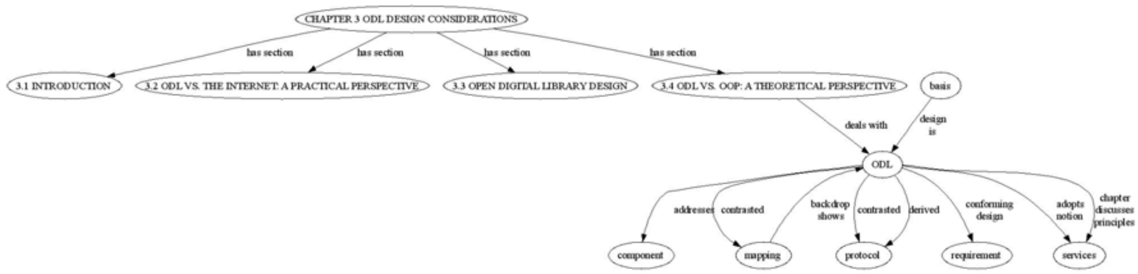


Figure 35: Automatically drawn concept map of an ETD. Information obtained from "Open Digital Libraries" by H. Suleman, chapter 3

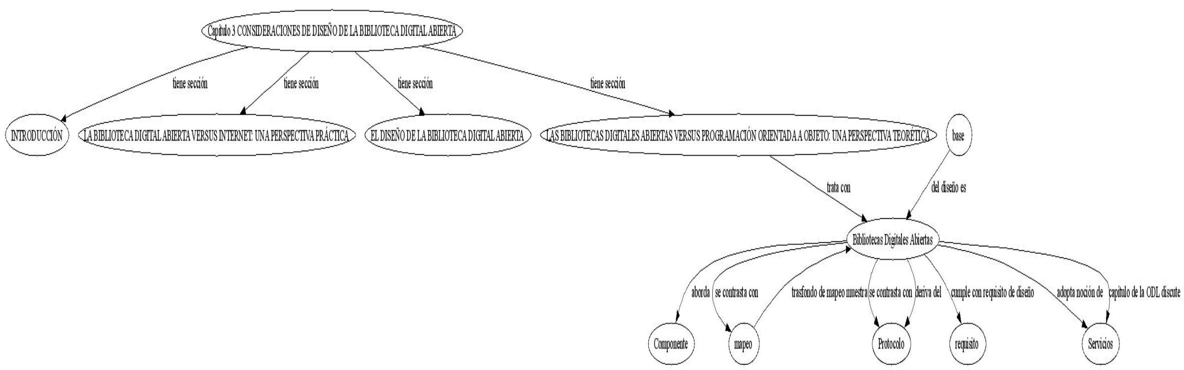


Figure 36: Automatically drawn concept map of an ETD, then translated into Spanish by a human. Information obtained from "Open Digital Libraries" by H. Suleman, chapter 3

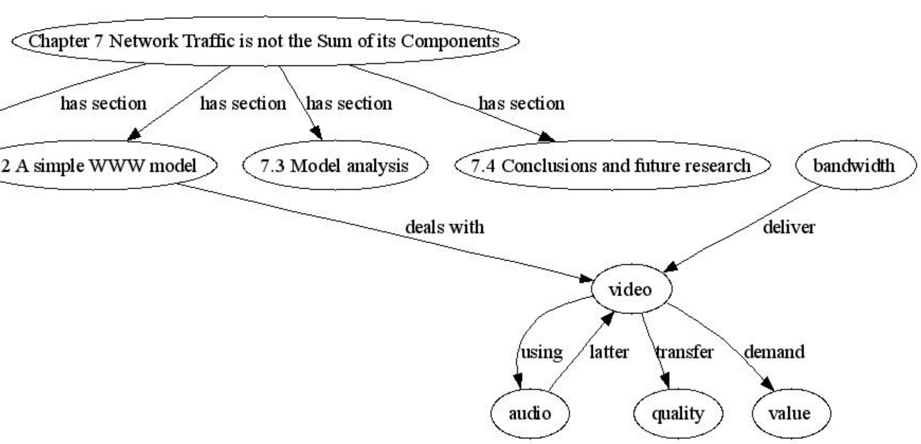


Figure 37: Automatically drawn concept map of an ETD. Information obtained from "Analysis and Modeling of World Wide Web Traffic" by Ghaleb Abdulla, chapter 7

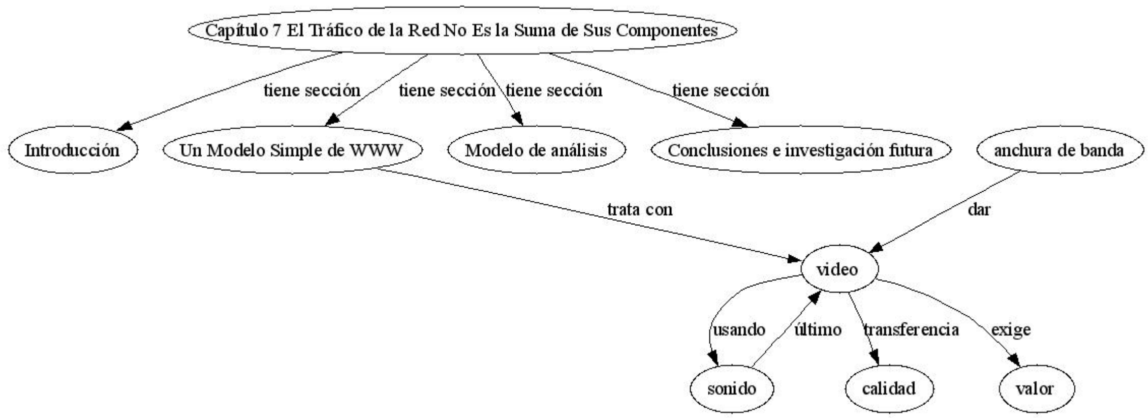


Figure 38: Automatically drawn concept map of an ETD, then machine-translated into Spanish. Information obtained from "Analysis and Modeling of World Wide Web Traffic" by G. Abdulla, chapter 7

Subjects were asked 6 questions about which of 5 dissertations are relevant to a particular question. A domain expert (a fellow graduate student in the Digital Libraries Research Laboratory at VT) came up with these questions, based on the 30 English ETDs, and listed which ETDs he considered relevant. Two more domain experts made their own relevance determinations about these dissertations for comparison purposes. The domain experts looked only at the English versions of the documents, plus the concept maps.

In order for the Mexican students to answer the questions, for each question they were provided one of 6 types of summaries of the dissertations.

7.1.2 Treatment conditions

1. Human translated abstract (A1)	4. Machine translated abstract (A2)
2. Human translated concept map (B1)	5. Machine translated concept map (B2)
3. Human translated abstract + Human translated concept map (C1)	6. Machine translated abstract + Machine translated concept map (C2)

Each subject was presented with 6 questions, one in each of the treatment conditions. For each question, they were allowed to pick from 5 dissertations. For instance, they were given a relevance question, and presented with 5 machine-translated concept maps (condition B2 above), based on 5 dissertations, and were asked which dissertations are relevant to that particular question based just on these concept maps.

Concerning the questions, each question had between 1 and 3 ETDs that are relevant to it. Each of the 6 questions always had the same 5 ETDs as possible answers. These 6 sets of ETDs are disjoint (6 * 5 = 30, making 30 total ETDs). Thus a subject never saw abstracts/concept maps of the same ETD for different questions. Each subject was presented with each ETD exactly once.

I randomized (using a computer program) the presentation order of the 6 questions and 6 treatment conditions into 7 possible presentation orders. Thus each presentation order was seen by about 3 subjects. These presentation orders are given in the table below, in the form (question number)-(treatment condition). These were randomized with the constraint that questions 1, 2, 3 always relate to treatment conditions X1 (X=A, B or C), while questions 4, 5, 6 always relate to condition X2. This is necessary because we did not have funding to produce human-translated abstracts and concept maps (conditions X1) for the other 15 dissertations.

Table 20: Presentation order for Experiment 5

	Order 1	Order 2	Order 3	Order 4	Order 5	Order 6	Order 7
First question presented to user	1C1	4B2	2B1	3B1	2A1	6A2	5C2
Second question	4B2	5C2	4B2	6C2	6C2	5B2	2B1
Third question	6C2	2C1	5A2	1C1	4A2	4C2	4A2
Fourth question	3B1	1B1	6C2	2A1	5B2	3C1	1A1
Fifth question	2A1	6A2	1A1	4A2	3B1	1A1	6B2
Sixth question	5A2	3A1	3C1	5B2	1C1	2B1	3C1

The experiment was designed to allow a within-subjects ANOVA on the results. All subjects saw 6 experimental conditions. The numbers compared in the ANOVA were number of agreements between the consensus of the 3 expert judgments and the subjects' judgments. I was careful to select question/ETD sets in which there was perfect inter-judge agreement on which ETDs were relevant and which were not. For each question, subjects were presented with 5 dissertations, so agreement could be either 0 (don't agree with judges on relevance of any dissertation), 1, 2, 3, 4, or 5 (agree with experts perfectly on relevance).

Since the English proficiency of the students could conceivably affect their answers, the subjects were given a pre-task question asking how many years they studied English, at the secondary school, university, and private language school levels. I also asked them to rate their English proficiency on a scale from 1 to 5, with the following levels:

Table 21: Spanish self-assessment levels

Proficiency Level (in Spanish)	**Translation**
1-No se inglés	I don't know English
2-Principiante	Beginner
3-Nivel Intermedio	Intermediate Level
4-Competente	Competent
5-Nivel Avanzado, entiende inglés de manera fluída	Advanced Level, understands English fluently

Further, I asked the students to list how many years they had studied CS, and whether they are familiar with digital libraries (the area to which most of the ETDs are related). I opted not to ask the subjects questions relating to their academic performance, such as their GPA or quartile, due to privacy issues.

Years of computer science study, years of English study, and self-assessment of English proficiency were treated as a covariates.

The English concept maps were translated into Spanish using several translation resources. For a given concept, my software first checks if the word or phrase occurs in the 47,000 entry U. of Maryland wordlist. The Maryland wordlist contains multiple possible Spanish translations for each English word/phrase, so I reordered the Maryland wordlist based on frequency in my Spanish ETD collection. For instance, both "web" and “tela” (spider-web) are possible Spanish translations of the English word “web”. Since “web” occurs more often in the Spanish computing ETD collection, occurrences of the English word “web” are translated as “web”, not “tela”. I supplemented the Maryland wordlist with translations from the ACM CC2001 classification scheme, prepared by Fernando Das Neves.

If an English phrase does not occur in the Maryland wordlist, my software checks in the list of phrases that I mined from the Spanish ETD collection using my implementation of the Lopez-Osteñero algorithm. If it is still not found, my software sends the word/phrase in question to Systran. It is unfortunate that I have to resort to Systran since it often gives poor results. For future work I would prefer to use a more-customizable machine translation system, tailored specifically to handle computing/technical terms.

7.1.3 Descriptive statistics

Twenty-two subjects responded to our call for participants. However, one subject answered only the first question, and left the other 9 blank. Therefore I removed his/her response from the data for analysis purposes. Thus, 21 subjects fully participated (N=21). The means for the agreement measure (ranging from 0 to 5) for the six possible combinations are shown in Table 22.

Table 22: Agreement between Subjects and Experts for 6 treatment conditions from 0 (no agreement) to 5 (perfect agreement).

	Abstract	Concept Map	Abstract + Concept Map
Human translated	3.19	4.00	3.71
Machine translated	3.05	3.90	3.48

7.1.4 ANOVA results

As stated before, I had 21 participants who answered all of the 6 relevance questions (N=21). The ANOVA revealed that the Presentation Type (levels were abstract, concept map, or both) had a significant effect on expert/subject agreement in relevance judgments. The Translation Type did not have a significant effect, nor did the Presentation Type * Translation Type interaction. See the ANOVA results below.

Tests of Within-Subjects Effects

Measure: Agreement

Table 23: ANOVA main effect results

Source		Type III Sum of Squares	df	F	Sig.
PresentationType	Sphericity Assumed	14.683	2	7.469	**.002**
	Greenhouse-Geisser	14.683	1.896	7.469	**.002**
	Huynh-Feldt	14.683	2.000	7.469	**.002**
	Lower-bound	14.683	1.000	7.469	**.013**
TranslationType	Sphericity Assumed	.794	1	2.016	.171
	Greenhouse-Geisser	.794	1.000	2.016	.171
	Huynh-Feldt	.794	1.000	2.016	.171
	Lower-bound	.794	1.000	2.016	.171
PresentationType * TranslationType	Sphericity Assumed	.111	2	.067	.935
	Greenhouse-Geisser	.111	1.967	.067	.933
	Huynh-Feldt	.111	2.000	.067	.935
	Lower-bound	.111	1.000	.067	.799

Computed using alpha = .05

I assume sphericity for the Presentation Type and Translation Type, since Mauchly's Test of Sphericity does not reject the null hypothesis. See below.

Table 24: Sphericity test results

Mauchly's Test of Sphericity(b)

Measure: Agreement

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound	Greenhouse-Geisser
PresentationType	.945	1.067	2	.587
TranslationType	1.000	.000	0	.
PresentationType * TranslationType	.983	.324	2	.851

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.
Within Subjects Design: PresentationType+TranslationType+PresentationType*TranslationType

7.1.5 Non-parametric tests

Since fewer than 30 subjects participated in the experiment, it is possible that a t-test could indicate that differences in means are significant even when they are not. Thus I tested the user agreement data with a non-parametric test. For related samples, the non-parametric equivalent of the t-test is the Wilcoxon signed-rank test [121]. The samples are related in this case because of the within-subjects design. Since the ANOVA results (see Table 23) did not identify translation type (human or machine) as significant, I combined the two translation types for analysis purposes.

Table 25 shows the means when both translation types are combined.

Table 25: Means of Presentation Types

Abstract	Concept Map	Abstract + CM
3.12	3.95	3.60

The table below indicates the p-values for the four comparisons of abstracts and concept maps.

Table 26: Results of Wilcoxon signed-rank tests

Test Statistics(c)

	ConceptMap - Abstract	AbstractPlusConceptMap - Abstract	AbstractPlusConceptMap - ConceptMap
Z	-3.363(a)	-2.216(a)	-1.668(b)
Asymp. Sig. (2-tailed)	.001	.027	.095

a Based on negative ranks.
b Based on positive ranks.
c Wilcoxon Signed Ranks Test

As one can see, both Concept Map and Abstract plus Concept Map have significantly higher means than Abstract alone. There was no significant difference between the Concept Map and the Abstract plus Concept Map condition.

7.1.6 Results of Likert questions

Users were asked to rate the concept maps on a Likert scale based on importance of nodes, relations between nodes, and overall usefulness of the maps, as well as the overall usefulness of the abstracts. The scale was 5 points, where 1=bad and 5=good. The results are below.

Table 27: Likert results for Experiment 5

User	Abstract	Concept Map Nodes	Concept Map Relations	Concept Maps Overall
1	1	4	4	4
2	3	3	3	4
3	2	4	5	4
4	2	5	3	5
5	3	5	4	5
6	2	4	3	4
7	2	4	4	5
8	5	3	4	3
9	3	4	4	4
10	2	4	2	4
11	5	4	4	4
12	3	3	4	4
13	2	4	3	4
14	2	5	4	4
15	3	4	4	4
16	5	4	4	5
17	3	4	4	4
18	3	4	3	4
19	4	3	3	4
20	4	3	3	5
21	5	4	4	5
Mean=	**3.05**	3.90	3.62	**4.24**
T-test				
p =	**0.00041**			

Subjects preferred the concept maps, overall, to the abstracts by a significant margin (*p=0.05*). Subjects did not find a significant difference between the quality of the nodes and that of the links. Subjects did however find the concept maps overall to be more helpful that either the nodes or links separately.

7.1.7 Results of covariate analyses

Subjects were asked how many years they had studied computer science. All 20 subjects who responded to this question (one left it blank) indicated they had studied computer science for 3, 4, or 5 years (mean = 3.86). Also, subjects were asked to rate their English language proficiency in two ways. First, they were asked how many years they have studied English. Secondly, they were asked to rate their proficiency on a scale from 1 to 5. The levels were labeled

- 1-No se inglés (I don't know English),
- 2-Principiante (Beginner),
- 3-Nivel Intermedio (Intermediate Level),
- 4-Competente (Competent), and
- 5-Nivel Avanzado, entiende inglés de manera fluída (Advanced, I understand English fluently).

All 21 subjects who responded to this questions rated their English as 3, 4, or 5 (mean=3.82). I compared the number of years they had studied English, to their self-assessment, and found that they were correlated at the *p=0.01* level. See Table 28 below.

Table 28: Correlation between years of studying English and self-assessment of English proficiency

Correlations

		YearsEnglish	SelfAssessEng
YearsEnglish	Pearson Correlation	1	**.649(**)**
	Sig. (2-tailed)		.002
	N	20	20
SelfAssessEng	Pearson Correlation	**.649(**)**	1
	Sig. (2-tailed)	.002	
	N	20	21

** Correlation is significant at the 0.01 level (2-tailed).

I treated Years of studying Computing Science, Years of studying English, and English Self-assessment as covariates with the main effects in the ANOVA. Unfortunately, one subject did not answer these pre-questions, so for this ANOVA, N=20 instead of N=21. The results are in Table 29.

Table 29: ANOVA results for Experiment 5

Tests of Within-Subjects Effects

Measure: Agreement

Source		Type III Sum of Squares	df	F	Sig.
PresentationType	Sphericity Assumed	1.501	2	.894	.419
	Greenhouse-Geisser	1.501	1.879	.894	.414
	Huynh-Feldt	1.501	2.000	.894	.419
	Lower-bound	1.501	1.000	.894	.358
PresentationType * YearsCS	Sphericity Assumed	.321	2	.191	.827
	Greenhouse-Geisser	.321	1.879	.191	.814
	Huynh-Feldt	.321	2.000	.191	.827
	Lower-bound	.321	1.000	.191	.668
PresentationType * YearsEnglish	Sphericity Assumed	7.554	2	4.501	**.019**
	Greenhouse-Geisser	7.554	1.879	4.501	**.021**
	Huynh-Feldt	7.554	2.000	4.501	**.019**
	Lower-bound	7.554	1.000	4.501	.050
PresentationType * SelfAssessEng	Sphericity Assumed	10.067	2	5.998	**.006**
	Greenhouse-Geisser	10.067	1.879	5.998	**.007**
	Huynh-Feldt	10.067	2.000	5.998	**.006**
	Lower-bound	10.067	1.000	5.998	**.026**
TranslationType	Sphericity Assumed	.008	1	.028	.870
	Greenhouse-Geisser	.008	1.000	.028	.870
	Huynh-Feldt	.008	1.000	.028	.870
	Lower-bound	.008	1.000	.028	.870
TranslationType * YearsCS	Sphericity Assumed	.011	1	.038	.848
	Greenhouse-Geisser	.011	1.000	.038	.848
	Huynh-Feldt	.011	1.000	.038	.848
	Lower-bound	.011	1.000	.038	.848
TranslationType * YearsEnglish	Sphericity Assumed	2.081	1	7.020	**.017**
	Greenhouse-Geisser	2.081	1.000	7.020	**.017**
	Huynh-Feldt	2.081	1.000	7.020	**.017**
	Lower-bound	2.081	1.000	7.020	**.017**
TranslationType * SelfAssessEng	Sphericity Assumed	.104	1	.350	.563
	Greenhouse-Geisser	.104	1.000	.350	.563
	Huynh-Feldt	.104	1.000	.350	.563
	Lower-bound	.104	1.000	.350	.563
PresentationType * TranslationType	Sphericity Assumed	1.099	2	.813	.452
	Greenhouse-Geisser	1.099	1.983	.813	.452
	Huynh-Feldt	1.099	2.000	.813	.452
	Lower-bound	1.099	1.000	.813	.381
PresentationType * TranslationType * YearsCS	Sphericity Assumed	.132	2	.098	.907

	Greenhouse-Geisser	.132	1.983	.098	.906
	Huynh-Feldt	.132	2.000	.098	.907
	Lower-bound	.132	1.000	.098	.759
PresentationType * TranslationType * YearsEnglish		4.390	2	3.249	.052
	Greenhouse-Geisser	4.390	1.983	3.249	.052
	Huynh-Feldt	4.390	2.000	3.249	.052
	Lower-bound	4.390	1.000	3.249	.090
PresentationType * TranslationType * SelfAssessEng	Sphericity Assumed	1.036	2	.767	.473
	Greenhouse-Geisser	1.036	1.983	.767	.472
	Huynh-Feldt	1.036	2.000	.767	.473
	Lower-bound	1.036	1.000	.767	.394

Computed using alpha = .05

As one can see, the PresentationType * YearsEnglish covariate relationship is significant (assuming sphericity), the PresentationType * SelfAssessEng relationship is significant, and the TranslationType * YearsEnglish relationship is significant at the *p=0.05* level. Interestingly, the TranslationType * SelfAssessEng relation is not significant. Years of studying CS (YearsCS) did not have a significant covariate effect with any of the main effects.

8 Conclusions

The results of Experiment 5 give us evidence that automatically-generated and translated concept maps can be effective indicative summaries of ETDs, with or without the use of automatically-translated abstracts. Surprisingly, the ANOVA results indicate no significant difference between the human translations and the machine translations, which supports my hypothesis that the short words and phrases that appear in concept maps can be translated more succinctly than the full text of abstracts. However, there is no doubt that the quality of the machine translation can be further improved. Experiment 5 did not test how much difference using the mined Spanish phrase translations made over simply using Systran translations. Incorporating the technical phrase translations into Systran (or some other off-the-shelf translation tool) would require considerable effort, whereas using the phrases to translate node labels in concept maps was very straightforward. This also supports my hypothesis that translating concept maps can be easier than translating abstracts in many cases.

8.1 Deliverables

Below are deliverables that I have produced in the course of my Ph.D. research and intend to make available to the research community.

1. Software to automatically generate concept maps from English documents using Relex plus the computer science-aware ontology that I have described. The code can be found at http://purl.oclc.org/NET/VTDLIB/CMAP/GENERATION/CODE.
2. The 200+ GATE-gazetteer lists that I have used to implement The Ontology Project's Computing Ontology.

3. Software to automatically translate English concept maps into Spanish, using code based on López-Ostenero's work [68], a U. of Maryland Spanish-English wordlist, and Fernando das Neves's translation of the ACM subject headings.
4. Due to licensing obligations, I cannot make the 438 ETDs from UNAM available to other researchers. I can however make the text of the 417 computing-related Spanish ETDs from Scirus, and the 89 ETDs from UDLA, publicly available. I also can make the list of 700,000+ Spanish computing-related phrases available for cross-language research. Assuming I can get permission from the U. of Maryland, I can make the slightly enlarged version of their Spanish/English word/phraselist available, including additions from the ACM subject headings.
5. I will make available a technical report describing results of experiment 4 and experiment 5, focusing on the cross-language aspect of experiment 5.

8.2 Future work

This study has given evidence that machine-translated concept maps are about as good in assisting students to make relevance determinations of ETDs as human translated concept maps are. Further, I still believe that there is much room for improvement in the quality of the machine-translated concept maps. In the case of concept maps generated by Relex, consisting only of terms found in a given ontology, one could translate the entire ontology once at the beginning of the process. As of March 20, 2007, the subset of The Ontology Project's computing ontology that I am using has only 7,663 words. Thus, it is much smaller than even a single ETD, which might have over 50,000 words. Translating the entire ontology at the beginning, despite being a fairly expensive task if done by professional translators, would make translating the concept maps much simpler from that point on. The translation process could be further simplified if one restricts the set of valid verb relations between the concepts to a reasonable number (say, less than 100), such that these could be translated *a priori* as well. The set of valid verbal relations would likely be specified by the ontology itself, although since The Ontology Project has not yet finished encoding their CS ontology into OWL, I have not included this is my work. The Ontology Project is currently working on defining the valid relations between various concepts in the computing ontology (Dr. Cassel, personal communication). These steps would make the automatic translation of the concept maps much quicker. It would be a simple lookup, with some grammatical rules to make minor changes to ensure the relationship remains readable. It would also improve the quality of the translations compared to a fully-automatic approach.

To find the translations of words and acronyms not found in our English/Spanish wordlist, I could take advantage of the similarities between English and Spanish and use spelling and cognate clues, as done by Koehn and Knight [57].

I plan to further investigate the effectiveness of providing links to the context of the original document, using techniques in addition to just hovertext. This could be done with the help of tools developed to support superimposed information [80]. I am investigating making our concept-map generation and translation scale to full collections, with the eventual goal of making the entire Virginia Tech CS-ETD collection, and perhaps collections from other universities, available in concept map form, including in other languages.

References

1. ACM. ACM Computing Classification Scheme, 1998,
2. ACM, IEEE-CS and AIS. Curriculum Recommendations of the ACM, IEEE-CS, AIS, ACM, 2005,
3. Agrawal, R., Imielinski, T. and Swami, A., Mining association rules between sets of items in large databases in *ACM SIGMOD Conference on Management of Data*, Washington, D.C., 1993, 207-216.
4. Agrawal, R. and Srikant, R., Fast Algorithms for Mining Association Rules in Large Databases in *20th International Conference on Very Large Databases (VLDB'94)*, Santiago, Chile, 1994, 487-499.
5. Amini, M. and Gallinari, P., The Use of Unlabeled Data to Improve Supervised Learning for Text Summarization in *SIGIR '02*, Tampere, Finland, 2002, ACM Press, New York.
6. Armitage, J., Dealing with constraints: "And another thing. The current site is in German": the final project in an international digital business consultancy in *2003 Conference on Designing for User Experiences*, San Francisco, 2003, ACM Press, New York, 1-15.
7. ASIIN. German Accreditation for Informatics Programs, ASIIN, Dusseldorf, 2002,
8. Ausubel, D.P. *Educational Psychology: A cognitive view*. Rinehart and Winston, New York, 1968.
9. Ausubel, D.P., Novak, J. D. & Hanesian, H *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston, New York, 1978.
10. Bahr, S. and Dansereau, D. Bilingual Knowledge maps (Bi-K Maps) in second language vocabulary learning. *Journal of Experimental Education*, *70* (1), 5-24.
11. Banerjee, S. and Pedersen, T., The Design, Implementation, and Use of the (N)gram (S)tatistic (P)ackage in *Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 2003, 370--381.
12. Borgelt, C., Efficient Implementations of Apriori and Eclat in *Workshop of Frequent Item Set Mining Implementations FIMI 2003*, Melbourne, FL, 2003, 26-34.
13. Borgelt, C., Recursion Pruning for the Apriori Algorithm in *Workshop of Frequent Item Set Mining Implementations FIMI 2004*, Brighton, UK, 2004.
14. Brockett, C., Aikawa, T., Aue, A., Menezes, A., Quirk, C. and Suzuki, H., English-Japanese Example-Based Machine Translation Using Abstract Linguistic Representations in *Workshop on Machine Translation in Asia*, Taiwan, 2002.
15. Brown, P.F., J., C., Pietra, S.A.D., Pietra, V.J.D., Jelinek, F., Lafferty, J., Mercer, R.L. and Roossin, P.S. A statistical approach to machine translation. *Computational Linguistics*, *16* (2), 79-85.
16. Buzydlowski, J.W., White, H.D. and Lin, X., Term Co-occurrence Analysis as an Interface for Digital Libraries in *Joint Conference on Digital Libraries, JCDL 2001*, Roanoke, Virginia, 2001, 133 - 144.

17. Callon, M. and Law, J. *Mapping the Dynamics of Science and Technology*. MacMillan, Basingstoke, UK, 1986.
18. Cañas, A.J., Hill, G., Carff, R., Suri, N., Lott, J., Gómez, G., Eskridge, T.C., Arroyo, M. and Carvajal, R., CMapTools: A Knowledge Modeling and Sharing Environment in *First Int. Conference on Concept Mapping*, Pamplona, Spain, 2004, Universidad Pública de Navarra, 16-24.
19. Cañas, A.J., Valerio, A., Lalinde-Pulido, J., Carvalho, M. and Arguedas, M., Using WordNet for Word Sense Disambiguation to Support Concept Map Construction in *International Symposium on String Processing and Information Retrieval*, Manaus, Brazil, 2003.
20. Carbonell, J. and Goldstein, J., The use of MMR, diversity-based reranking for reordering documents and producing summaries in *SIGIR*, Melbourne, 1998, ACM Press, New York, 335-336.
21. Carnot, M.J., Dunn, B., Cañas, A.J., Gram, P. and Muldoon, J. Concept Maps vs. Web Pages for Information Searching and Browsing, Institute for Human and Machine Cognition, Pensacola, FL, 2001,
22. Cassel, L.N. Information organization for the union of computing related disciplines, 2005,
23. Cassel, L.N. The Ontology Project. Cassel, L.N. ed., Villanova University, Philadelphia, 2006,
24. Chaffin, R. and Herrmann, D.J. Relation Element Theory: A New Account of the Representation and Processing of Semantic Relations. in Gorfein, D.S. and Hoffman, R.R. eds. *Memory and Learning*, Lawrence Erlbaum, Hillsdale, NJ, 1987, 221-245.
25. Chmielewski, T.L. and Dansereau, D.F. Enhancing the recall of text: Knowledge mapping training promotes implicit transfer. *Journal of Educational Psychology*, *90* (3), 407-413.
26. Coffey, J.W., Carnot, M.J., Feltovich, P.J., Hoffman, R.R., Cañas, A.J. and Novak, J.D. A Summary of Literature Pertaining to the Use of Concept Mapping Techniques and Technologies for Education and Performance Support *Technical Report submitted to the US Navy Chief of Naval Education and Training*, Chief of Naval Education and Training, Pensacola, FL, 2003,
27. Cooper, J., Visualization of Relational Text Information for Biomedical Knowledge Discovery in *Information Visualization Interfaces for Retrieval and Analysis workshop, presented at JCDL 2003*, Houston, Texas, 2003.
28. Corney, D.P.A., Buxton, B.F., Langdon, W.B. and Jones, D.T. BioRAT: Extracting Biological Information from Full-Length Papers. *Bioinformatics*, *20* (17), 3206-3213.
29. Cummings, A.L., Halberg, E.T., Martin, J., Slemon, A.G. and Hiebert, B. Implications of counselor conceptualizations for counselor education. *Counselor Education and Supervision*, *30*, 120-134.
30. Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V., Framework and Graphical Development Environment for Robust NLP Tools and Applications in *40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, Philadelphia, 2002.

31. **Dejean, H., Gaussier, E. and Sadat, F., An approach based on multilingual thesauri and model combination for bilingual lexicon extraction in *19th International Conference on Computational Linguistics*, 2002, 1-7.**
32. **Dunning, T. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, *19* (1), 61-74.**
33. **Esiobu, G. and Soyibo, K. Effects of concept and vee mapping under three learning modes on students' cognitive achievement in ecology and genetics. *Journal of Research in Science Teaching*, *32* (9), 971-995.**
34. **Fellbaum, C. *WordNet: An electronic lexical database*. The MIT Press, Cambridge, Massachusetts, 1998.**
35. **Friedman GENIES: A natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, *17* (Suppl 1), 74-82.**
36. **Friedman, C., Hripcsak, G., DuMouchel, W., Johnson, S.B. and Clayton, P.D. Natural Language processing in an operational clinical information system. *Natural Language Engineering*, *1* (1), 83-108.**
37. **Fung, P., A Pattern Matching Method for Finding Noun and Proper Noun Translations from Noisy Parallel Corpora in *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics.*, 1995.**
38. **Gaines, B.R. and Shaw, M.L.G., Using knowledge acquisition and representation tools to support scientific communities in *AAAI'94: Proceedings of the Twelfth National Conference on Artificial Intelligence*, Menlo Park, California, 1994, AAAI Press/MIT Press, 707-714.**
39. **Gansner, E.R. and North, S.C. An Open Graph Visualization System and Its Applications to Software Engineering. *Software: Practice and Experience*, *30* (11), 1203-1233.**
40. **Garber, S. Birds of a Feather? How Politics and Culture Affected the Designs of the U.S. Space Shuttle and the Soviet Buran *Science, Technology and Public Policy*, Virginia Tech, Blacksburg, Virginia, 2002, 64,**
41. **Gaussier, E., Renders, J., Matveeva, I., Goutte, C. and Dejean, H., A Geometric View on Bilingual Lexicon Extraction from Comparable Corpora in *ACL 2004*, Barcelona, 2004, 526-533.**
42. **Gey, F., Kando, N. and Peters, C., Cross Language Information Retrieval: A Research Roadmap in *Workshop at SIGIR-2002 International Conference On Research and Development in Information Retrieval*, Tampere, Finland, 2002, 1-9.**
43. **Gopan, G.D. and Swan, J.A. The science of scientific writing. *American Scientist*, *78*, 550-558.**
44. **Gordon, S.E. and Gill, R.T. Question probes: A structured method for eliciting declarative knowledge. *AI Applications in Natural Resource Management*, *3*, 13-20.**
45. **Gordon, S.E., Schmierer, K.A. and Gill, R.T. Conceptual graph analysis: Knowledge acquisition for instructional system design. *Human Factors*, *35*, 459-481.**

46. Hall, R., Dansereau, D. and Skaggs, L. Knowledge maps and the presentation of related information domains. *Journal of Experimental Education*, *61* (1), 5-18.
47. Honrado, A., Leon, R., O'Donnel, R. and Sinclair, D., A Word Stemming Algorithm for the Spanish Language in *International Symposium on String Processing Information Retrieval (SPIRE'00)*, Coruña, Spain, 2000, 139.
48. Hovy, E. and Lin, C.Y. Automatic Text Summarization in SUMMARIST. in Maybury, M.T. ed. *Advances in Automatic Text Summarization*, MIT Press, Cambridge, 1999, 81-94.
49. Hull, D., A practical approach to terminology alignment in *First Workshop on Computational Terminology (Computerm)*, Montreal, 1998, 1-7.
50. IBM. Unstructured Information Management Architecture (UIMA), 2006,
51. Intelligent_Software_Lab. POSBIOTM/W: POSTECH Bio-Text Mining System Workbench, 2005, http://isoft.postech.ac.kr/Research/Bio/bio.html.
52. Jones, S. and Paynter, G.W., Human evaluation of Kea, an automatic keyphrasing system in *First ACM/IEEE Joint Conference on Digital Libraries*, Roanoke, Virginia, 2001, ACM Press, 148-156.
53. Kaji, H. and Morimoto, Y., Unsupervised Word Sense Disambiguation Using Bilingual Comparable Corpora in *International Conference on Computational Linguistics 2002*, Taipei, 2002.
54. Kan, M.Y., McKeown, K.R. and Klavans, J.L., Applying natural language generation to indicative summarization in *8th European workshop on Natural Language Generation*, Toulouse, France, 2001, Association for Computational Linguistics, Morristown, NJ, 1-9.
55. Kitani, T. and Mitamura, T. An accurate morphological analysis and proper name identification for Japanese text processing. *Journal of Information Processing Society of Japan*, *35* (3), 404-413.
56. Klavans, J. and Tzoukermann, E., The BICORD System in *13th Conference on Computational Linguistics*, 1990, 174-179.
57. Koehn, P. and Knight, K., Learning a Translation Lexicon from Monolingual Corpora in *ACL Workshop on Unsupervised Lexical Acquisition*, Philadelphia, 2002.
58. Kohonen, T. The self-organizing Map. *Proceedings of the IEEE*, *78* (9), 1464-1480.
59. Kraaij, W., Nie, J. and Simard, M. Embedding web-based statistical translation models in cross-language information retrieval. *Computational Linguistics*, *29* (3), 381-419.
60. Lambiotte, J. and Skaggs, L.D., D. Learning from lectures: Effects of knowledge maps and cooperative review strategies. *Applied Cognitive Psychology*, *7*, 483-497.
61. Lavoie, B., White, M. and Korelsky, T., Learning domain-specific transfer rules: an experiment with Korean to English translation in *International Conference on Computational Linguistics*, Taipei, 2002, ACM, 1-7.
62. Leed, R.L. and Nakhimovsky, A.D. Lexical Functions and Language Learning. *Slavic and East European Journal*, *23* (1), 104-113.

63. Levin, R. and Redell, D.D. An Evaluation of the Ninth SOSP Submissions or How and How Not to Write a Good Systems Paper. *ACM SIGOPS Operating Systems Review*, *17* (3), 35-40.
64. Lin, D. and Pantel, P. Discovery of Inference Rules for Question Answering. *Natural Language Engineering*, *7* (4), 343-360.
65. Liu, H. MontyTagger, Cambridge, Mass, 2003, http://web.media.mit.edu/~hugo/montytagger/.
66. Liu, X. The Chemistry of Chinese Language, Twente University Press, Twente, Netherlands, 2002, 1-304,
67. Loh, S., Wives, L.K. and Oliveira, J.P.M. Concept-based Knowledge Discovery in Texts Extracted from the Web. *ACM SIGKDD Explorations*, *2* (1), 29-39.
68. López-Ostenero, F., Gonzalo, J. and Verdejo, F. Noun Phrases as building blocks for Cross-Language Search Assistance. *Information Process and Management*, *41*, 549-568.
69. Luhn, H.P. The Automatic Creation of Literature Abstracts. in Maybury, M.T. ed. *Advances in Text Summarization*, The MIT Press, Cambridge, 1958, 15-22.
70. Malone, J. and Dekkers, J. The concept map as an aid to instruction in science and mathematics. *School Science and Mathematics*, *84* (3), 220-231.
71. Marshall, B., Zhang, Y., Chen, H., Lally, A., Shen, R., Fox, E. and Cassel, L., Convergence of knowledge management and E-Learning: the GetSmart experience in *Joint Conference on Digital Libraries*, Houston, Texas, 2003, IEEE Press.
72. Mattox, D., Seligman, L. and Smith, K., Rapper: a wrapper generator with linguistic knowledge in *ACM Workshop on Information and Data Management*, Kansas City, 1999, ACM Press, 6-11.
73. Mauldin, M.L. Information Retrieval by Text Skimming *School of Computer Science*, Carnegie Mellon University, Pittsburgh, 1989, 1-179,
74. Mayfield, W.A., Kardash, C.M. and Kivlighan, D.M. Differences in experienced and novice knowledge structures about clients: Implications for case conceptualization. *Journal of Counseling Psychology*, *46*, 504-514.
75. McDonald, D. and Chen, H., Using sentence-selection heuristics to rank text segments in TXTTRACTOR in *International Conference in Digital Libraries*, Portland, OR, 2002, ACM Press, 28-35.
76. McNaught, C. and Kennedy, D. Use of Concept Mapping in the design of learning tools for interactive multimedia. *Journal of Interactive Learning Research*, *8* (3-4), 389-406.
77. McNeese, M.D., Zaff, B.S., Peio, K.J., Snyder, D.E., Duncan, J.C. and McFarren, M.R. An advanced knowledge and design acquisition methodology: Application for the Pilot's Associate., Human Systems Division, Aerospace Medical Research Laboratory, Air Force Systems Command, Wright-Patterson AFB, Ohio, 1990,
78. Mintzes, J.J., Wandersee, J.H. and Novak, J.D. *Assessing science understanding: A human constructivist view*. Academic Press, San Diego, 2000.

79. Muller, H., Kenny, E.E. and Sternberg, P.W. Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. *PLOS Biology*, *2* (11).
80. Murthy, U., Richardson, R., Fox, E.A. and Delcambre, L., Enhancing Concept Mapping Tools Below and Above to Facilitate the Use of Superimposed Information in *Concept Mapping Conference 2006*, San Jose, Costa Rica, 2006, University of Costa Rica, 197-203.
81. Nagao, M., A framework of a mechanical translation between Japanese and English by analogy principle. in *NATO symposium on Artificial and Human Intelligence*, Lyon, France, 1984, NATO Publications, 173-180.
82. Nagao, M. and Sato, S., Toward memory-based translation in *COLING 1990*, Helsinki, Finland, 1990, 247-252.
83. NDLTD. Networked Digital Library of Theses and Dissertations, 2005,
84. Nguyen, I.X., Xu, H., Yuan and Chen, B., INSPIRE: An Interactive Image Assisted Non-Photorealistic Rendering System in *Pacific Graphics 03*, Canmore, Canada, 2003, IEEE Computer Society Press, 472.
85. Nilsson, R.M. and Mayer, R.E. The effects of graphical organizers giving cues to the structure of a hypertext document on users' navigation strategies and performance. *International Journal of Human-Computer Studies*, *57* (1), 1-26.
86. Novak, J.D. *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in schools and Corporations.* Lawrence Erlbaum Associates, Mahwah, New Jersey, 1998.
87. Novak, J.D. and Gowin, D.B. *Learning How To Learn*. Cambridge University Press, Cambridge, UK, 1984.
88. Oard, D. Spanish/English text-mined wordlist. Richardson, R. ed., University of Maryland, College Park, MD, 2005,
89. Pantel, P. and Lin, D., Discovering Word Sense from Text in *Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Alberta, Canada, 2002, ACM Press, 613-619.
90. Paynter, G.W., Witten, I.H. and Nevill-Manning, C.G., Domain-specific keyphrase extracion in *Sixteenth International Joint Conference on Artificial Intelligence*, San Franciso, CA, 1999, Morgan Kaufmann Publishers, 668-673.
91. Rajaraman, K. and Tan, A., Knowledge Discovery from Texts: A Concept Frame Graph Approach in *International Conference on Information and Knowledge Management (CIKM'02)*, McLean, VA, 2002, 669 - 671.
92. Rau, L.F., Extracting company names from text in *Sixth IEEE Conference on Artificial Intelligence Applications*, 1991, 29-32.
93. Rewey, K., Dansereau, D., Dees, S. and Skaggs, L. Scripted cooperation and knowledge map supplements: Effects on the recall of biological and statistical information. *Journal of Experimental Education*, *60* (2), 93-107.
94. Rewey, K., Dansereau, D., Skaggs, L. and Hall, R. Effects of scripted cooperation and knowledge maps on the processing of technical material. *Journal of Educational Psychology*, *81* (4), 604-609.

95. Rex, D.B., Risch, J.S., Dowson, S.T. and Moon, B.D., Multiple source information analysis, GIS and Starlight in *7th ACM international symposium on Advances in geographic information systems*, Kansas City, Missouri, 1999, 158-159.
96. Rindflesch, T., Tanabe, L., Weinstein, J. and Hunter, L., EDGAR: extraction of drugs, genes and relations from the biomedical literature in *Pacific Symposium on Biocomputing 5*, Honolulu, 2000, 514-525.
97. Risch, J.S., Rex, D.B., Dowson, S.T., Walters, T.B., May, R.A. and Moon, B.D., The STARLIGHT Information Visualization System in *IEEE International Conference on Information Visualization*, London, 1997, 42-49.
98. Roth, W. and Roychoudhury, A. The concept maps as a tool for the collaborative construction of knowledge: A microanalysis of high school physics students. *Journal of Research in Science Teaching*, *30* (5), 503-554.
99. Rubin, D.L., Lewis, S.E., Mungall, C.J., Misra, S., Westerfield, M., Ashburner, M., Sim, I., Chute, C.G., Solbrig, H., Storey, M., Smith, B., Day-Richter, J., Noy, N.F. and Musen, M.A. National Center for Biomedical Ontology: Advancing Biomedicine through Structured Organization of Scientific Knowledge. *OMICS: A Journal of Integrative Biology*, *10* (2), 185-198.
100. Ruiz-Primo, M.A. and Shavelson, R.J. Problems and issues in the use of Concept Maps in science assessment. *Journal of Research in Science Teaching*, *33* (6), 569-600.
101. Safayeni, F., Derbentseva, N. and Cañas, A.J. Concept Maps: A Theoretical Note on Concepts and the Need for Cyclic Concept Maps, 2003,
102. Saito, H., Ohno, T., Ozaki, F., Saito, K., Maeda, T. and Ohuchi, A., A Semi-Automatic Construction Method of Concept Map Based on Dialog Contents in *Sixth International Conference on Computers in Education (ICCE2001)*, Seoul, South Korea, 2001, 31-38.
103. Settles, B. ABNER: an open source tool for automatically tagging genes, proteins, and other entity names in text. *Bioinformatics*, *21* (14), 3191-3192.
104. Shackelford, R. Why can't smart people figure out what to do about computing education? 2005, http://www.iste.org/Content/NavigationMenu/Professional_Development/Programming_at_NECC/Symposia/Computer_Science/2005/February2/Speaker_Presentations/shackelford-computing-ed.pdf.
105. Shaka, F. and Bitner, B., Construction and validation of a rubric for scoring concept maps in *Association for the Education of Teachers of Science Conference*, Pennsylvania State University, 1996, 650-669.
106. Sleator, D. and Temperley, D., Parsing English with a link grammar in *Third International Workshop on Parsing Technologies*, Tilburg, Netherlands & Durbuy, Belgium, 1993.
107. Society, A.C. Australian Computer Society List of Computing Terminology, Australian Computer Society, 2006,
108. Sowa, J.F. Semantic Networks. in Shapiro, S.C. ed. *Encyclopedia of Artificial Intelligence*, John Wiley & Sons, Inc., New York, 1992.

109. **Sparck-Jones, K. *Automatic Keyword Classification for Information Retrieval*. Archon Books, London, 1971.**
110. **Sparck-Jones, K. Discourse modeling for automatic summarizing, University of Cambridge Computer Laboratory, Cambridge, 1993,**
111. **Stefanov, K., Computing Ontology Creation in *CompSysTech'2003 International Conference*, Sofia, Bulgaria, 2003, 19-20.**
112. **Strzalkowski, T., Wang, J. and Wise, B., A robust practical text summarization system in *Proceedings of the Fifteenth National Conference on AI*, 1998, 26-30.**
113. **Sundheim, B. (ed.), *Proceedings of the Third Message Understanding Evaluation and Conference*. Morgan Kaufmann, Los Altos, CA, 1991.**
114. **Thorsden, M.L., A comparison of two tools for cognitive task analysis: concept Mapping and the Critical Decision Method in *Human Factors and Ergonomics Society 35th Annual Meeting*, Santa Monica, CA, 1991, Human Factors and Ergonomics Society, 283-285.**
115. **Wallace, J. and Mintzes, J.J. The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching*, *27* (10), 1033-1052.**
116. **Weideman, M. and Kritzinger, W., Concept mapping vs. web page hyperlinks as an information retrieval interface: preferences of postgraduate culturally diverse learners in *Annual research conference of the South American institute of computer scientists and information technologists on Enablement through technology*, Pretoria, South Africa, 2003, South African Institute for Computer Scientists and Information Technologists, 69-82.**
117. **White, R. and Gunstone, R. *Probing understanding*. Falmer Press, New York, 1992.**
118. **White, R.W., Song, H. and Liu, J., Concept Maps to Support Oral History Search and Use in *Joint Conference on Digital Libraries*, Chapel Hill, North Carolina, 2006, ACM Press, 192-193.**
119. **Wiegmann, D.A., Dansereau, D.F., McCagg, E.C., Rewey, K.L. and Pitre, U. Effects of Knowledge Map Characteristics on Information- Processing. *Contemporary Educational Psychology*, *17* (2), 136-155.**
120. **Wierzbicka, A. *Semantics, Primes and Universals*. Oxford University Press, Oxford, UK, 2006.**
121. **Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.**
122. **Willerman, M. and MacHarg, R.A. The concept map as an advance organizer. *Journal of Research in Science Teaching*, *28* (8), 705-711.**
123. **Wong, P.C., Whitney, P. and Thomas, J., Visualizing Association Rules for Text Mining in *IEEE Information Visualization '99*, Los Alamitos, CA, 1999, IEEE CS Press, 120-123.**
124. **Wu, D., Grammarless Extraction of Phrasal Translation Examples from Parallel Texts in *Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, 1995.**

125. Zaki, M.J. and Phoophakdee, B. MIRAGE: A Framework for Mining, Exploring and Visualizing Association Rules, Rensselaer Polytechnic Institute, Troy, NY, 2000, 1-23, http://www.cs.rpi.edu/research/pdf/03-04.pdf.

126. Zhang, L. Knowledge Graph Theory and Structural Parsing, Twente University, Twente, Netherlands, 2002, 1-234,