# Towards A Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning

Raymond Lau, *Member, IEEE,* Dawei Song, Yuefeng Li *Member, IEEE,* Terence Cheung *Member, IEEE,* and Jin-Xing Hao

***Abstract*— With the wide spread applications of e-Learning technologies to education at all levels, increasing number of online educational resources and messages are generated from the corresponding e-Learning environments. Accordingly, instructors are often overwhelmed by the huge number of messages created by students through online discussion forums. It is quite difficult, if not totally impossible, for instructors to read through and analyze these messages to understand the progress of their students on the fly. As a result, adaptive teaching for a large class is handicapped. The main contribution of this paper is the illustration of a novel concept map generation mechanism which is underpinned by a fuzzy domain ontology extraction algorithm. The proposed mechanism can automatically construct concept maps based on the messages posted to online discussion forums. Our initial experimental results reveal that the accuracy and the quality of the automatically generated concept maps are promising. Our research work opens the door to the development and application of intelligent software tools to enhance e-Learning. To our best knowledge, the work presented in this paper demonstrates the first application of fuzzy domain ontology extraction method to facilitate adaptive e-Learning.**

***Index Terms*— Domain Ontology, Ontology Extraction, Text Mining, Fuzzy Sets, Concept Map, e-Learning.**

## I. Introduction

Electronic learning (e-Learning) refers to the application of information and communication technologies (e.g., Internet, multimedia, etc.) to enhance ordinary classroom teaching and learning [2], [51]. With the maturity of the technologies such as the Internet and the decreasing cost of the hardware platforms, more institutions are adopting e-Learning as a supplement to traditional instructional methods [49], [51]. In fact, one of the main advantages of e-Learning technology is that it can facilitate *adaptive learning* such that instructors can dynamically revise and deliver instructional materials in accordance with learners' current progress. In general, adaptive teaching and learning refers to the use of what is known about learners, a priori or through interactions, to alter how a learning experience unfolds, with the aim of improving learners' success and satisfaction [16]. The current state-of-the-art of e-Learning technology supports automatic collection of learners' performance data (e.g., via online quiz) [13]. However, few of the existing e-Learning technologies can support automatic analysis of learners' progress in terms of the knowledge structures they have acquired. In this paper, we illustrate a methodology of automatically constructing concept maps to characterize learners' understanding for a particular topic; thereby instructors can conduct adaptive teaching and learning based on the learners' knowledge structures as reflected on the concept maps. In particular, our concept map generation mechanism is underpinned by a context-sensitive text mining method [24] and a fuzzy domain ontology extraction algorithm.

The notion of ontology is becoming very useful in various fields such as intelligent information extraction and retrieval, semantic Web, electronic commerce, and knowledge management [34], [56]. Although there is not a universal consensus on the precise definition of ontology, it is generally accepted that ontology is a formal specification of conceptualization [14]. Ontology can take the simple form of a taxonomy of concepts (i.e., light weight ontology), or the more comprehensive representation of comprising a taxonomy, as well as the axioms and constraints which characterize some prominent features of the real-world (i.e., heavy weight ontology) [5]. Domain ontology is one kind of ontology which is used to represent the knowledge for a particular type of application domain [10]. On the other hand, concept maps are used to elicit and represent the knowledge structure such as concepts and propositions as perceived by individuals [37]. Concept maps are similar to ontology in the sense that both of these tools are used to represent concepts and the semantic relationships among concepts. However, ontology is a formal knowledge representation method to facilitate human and computer interactions and it can be expressed by using formal semantic markup languages such as RDF and OWL [9], whereas concept map is an informal tool for humans to specify semantic knowledge structure. Figure 1 shows an example of the owl statements describing one of the fuzzy domain ontologies automatically generated from our system. It should be noted that we use the (rel) attribute of the `<rdfs:comment>` tag to describe the membership of a fuzzy relation (e.g., the super-class/sub-class relationship). We only focus on the automatic extraction of lightweight domain ontology in this paper. More specifically, the lightweight fuzzy domain ontology is used to generate concept maps to represent learners' knowledge structures.

### A. The Problem Area

With the rapid growth of the applications of e-Learning to enhance traditional instructional methods, it is not surprising

R. Lau, T. Cheung, and J. Hao are with the Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR. E-mail:raylau@cityu.edu.hk.

D. Song is with the Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, U. K. E-mail:d.song@open.ac.uk.

Y. Li is with the Faculty of Information Technology, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. E-mail:y2.li@qut.edu.au

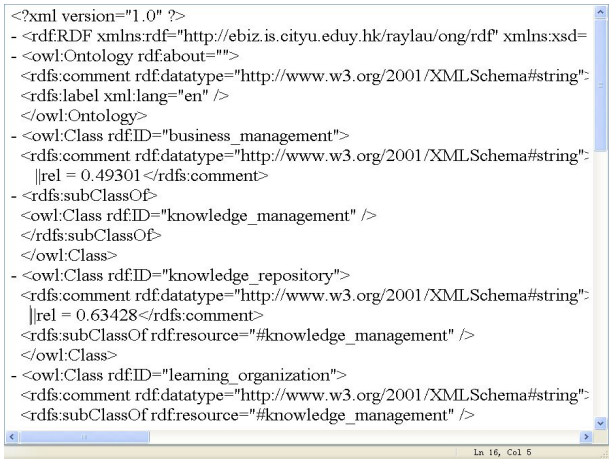


Fig. 1. A Segment of OWL Statements for a Knowledge Management Ontology

to find that there are new issues or challenges arising when educational practitioners try to bring information technologies down to their classrooms [3], [13]. The situation is similar to the phenomenon of the rapid growth of the Internet and the World Wide Web (Web). The explosive growth of the Web makes information seekers become increasingly more difficult to find relevant information they really need. This is the so-called problem of information overload [28]. With respect to e-learning, the increasing number of educational resources deployed online and the huge number of messages generated from online interactive learning (e.g., Blogs, emails, chat rooms) also lead to the excessive information load on both the learners and the instructors. For example, to promote reflexive and interactive learning, instructors often encourage their students to use online discussion boards, blogs, or chat rooms to reflect what they have learnt and to share their knowledge with other fellow students during or after normal class time. With the current practice, instructors need to read through all the messages in order to identify the actual progress of their students. From the pedagogical point of view, such an analysis process is essential since instructors have to understand the cognitive states of their students in order to conduct adaptive teaching and learning. Nevertheless, manually browsing and analyzing the huge number of messages is very time-consuming, and it is extremely difficult, if not totally impossible, to conduct the analysis process in the middle of a lecture or a tutorial session.

### *B. The Proposed Approach*

To alleviate the excessive information overload imposed on instructors and to facilitate adaptive teaching, we develop an automated concept map generation tool to assist instructors to analyze and visualize the latent knowledge structures embedded in the large number of messages posted to online discussion forums, blogs, or chat rooms. Previous research has indicated that an individual's problem solving performance in knowledge-intensive domains could be predicted by analyzing the structural and/or content properties of the concept map developed by the individual [37]. Studies on problem-solving and reasoning ability also demonstrate that successful learners can develop elaborately complex and highly integrated structures of related concepts [19], [41]. Based on these observations, a context-sensitive text mining method, which combines lexico-syntactic and statistical learning approaches, is applied to extract prominent concepts from the online messages entered into an e-Learning platform [23]. In addition, a subsumption based fuzzy domain ontology extraction algorithm is applied to infer the taxonomy structure from the set of prominent concepts. The resulting fuzzy domain ontology is then used to generate the corresponding concept maps which disclose the knowledge structure acquired by an individual or a group of learners. Thereby, instructors can quickly and easily observe the progress of their students to conduct adaptive teaching on the fly.

### *C. The Contributions*

The main contributions of our research work are two folds; from the theoretical stand point, we contribute to the development of a novel fuzzy domain ontology extraction method which alleviates with the knowledge acquisition bottle-neck of manually constructing domain ontologies. Although some learning techniques have been proposed for automatic or semi-automatic extraction of domain ontology before [4], [10], [45], these methods are still primitive and further enhancement in terms of computational efficiency and learning accuracy is required. Since ontology extraction from text often involves uncertainty (e.g., which messages (objects) are associated with which concepts (classes)), an uncertainty representation and management mechanism is required to address such an issue. It is believed that the notions of fuzzy set and fuzzy Relation provide a sound and rigorous method to represent knowledge with uncertainty [60]. One of our contributions is the development of a formal fuzzy domain ontology model which is underpinned by fuzzy sets and fuzzy relations. Moreover, based on the concept of subsumption, we have developed a fuzzy domain ontology extraction algorithm for the automatic extraction of domain ontologies from text.

From the practical stand point, our research work opens the door to the development of intelligent software tools for enhancing e-Learning technology. In particular, we have demonstrated how to apply the context-sensitive text mining method and the fuzzy domain ontology extraction algorithm to automatically generate concept maps to reveal the knowledge structures of students who are engaging in e-Learning. As a result, instructors can conduct adaptive teaching and learning based on the information disclosed on the concept maps.

### *D. Outline of the Paper*

The remainder of the paper is organized as follows. Section II highlights previous research in the related area and compare these research work with ours. A framework of fuzzy domain ontology based concept map generation is highlighted in Section III. The cognitive and linguistic foundations of the proposed context-sensitive text mining method for concept extraction are illustrated in Section IV. Then, the computational algorithm of the fuzzy domain ontology extraction method is depicted in Section V. Section VI explains how

the fuzzy domain ontology extraction method is applied to adaptive e-Learning. Section VII describes the evaluation of the proposed fuzzy domain ontology based concept map generation mechanism. Finally, we offer concluding remarks and describe future direction of our research work.

## II. Related Research

There is a large number of educational intermediaries storing meta-data descriptions for various learning resources to facilitate educational knowledge management [38]. In order to ensure effective communications between the users and the learning resources, automatic discovery of the taxonomies of these learning resources is required. A data mining approach was proposed to discover the relations of the meta-data describing the various learning resources. Terms from the meta-data description files were parsed and stop words were removed. Language engineering tools such as WordNet [33] was applied to extract the word roots (lemmatization) and the Brill tagger algorithm was used for part of speech tagging. As a result, a set of unique keywords could be extracted. A data matrix with each column corresponding to a learning resource and each row corresponding to a keyword was developed. A graph-based clustering algorithm was then applied to the data matrix to extract meaningful concepts for the learning resources and to identify the relations among the concepts. Our work aims at extracting and visualizing the concept maps based on the online messages created by the learners rather than discovering the ontology of educational resources. We employ a hybrid lexico-syntactic and statistical learning method rather than a computationally expensive graph-based approach for ontology extraction. Moreover, we employ the notion of fuzzy ontology rather than crisp ontology to explicitly model the uncertainty arising in automated ontology extraction.

There was also research work exploring the ideas of automatically extracting ontologies from teaching documents although the algorithmic details were not illustrated [20]. Previous work had also employed the Term Frequency Inverse Document Frequency (TFIDF) heuristic developed from the field of IR to extract prominent concepts from electronic messages generated in e-Learning [58]. A knowledge density score was developed based on the TFIDF term weighting formula to assess the extent of contribution to online knowledge sharing by individuals. Our document parsing approach also employs TFIDF and other linguistic pattern recognition method to extract concepts from text. In addition, we deal with the automatic construction of a taxonomy of concepts as well.

The FOGA framework for fuzzy ontology extraction has been reported [53]. The FOGA framework consists of fuzzy formal concept analysis, fuzzy conceptual clustering, fuzzy ontology generation, and semantic representation conversion. Essentially, the FOGA method extends the formal concept analysis approach with the notions of fuzzy sets. The notions of formal context and formal concept have been fuzzified by introducing the respective membership functions. In addition, an approximate reasoning method is developed so that the automatically generated fuzzy ontology can be incrementally furnished with the arrival of new instances. The FOGA framework is evaluated in a small citation database. Our method discussed in this paper differs from the FOGA framework in that a more compact representation of fuzzy domain ontology is developed. Our proposed method is based on the previous work in computational linguistic and with the computational algorithm developed with respect to the concept of fuzzy relations. We believe that the proposed method is computationally more efficient and be able to scale up for the huge textual databases which typically consists of millions of records and thousands of terms. Finally, our proposed method is validated in a standard benchmark textual database which is considerably larger than the citation database used in [53].

A fuzzy ontology which is an extension of the crisp domain ontology was utilized for news summarization purpose [27]. In this semi-automatic ontology discovery approach, the domain ontology about various events covered by some net news was manually developed by human domain experts. A document pre-processing mechanism extracted the meaningful terms from news corpus with the help of a Chinese news dictionary created by the domain experts. The meaningful terms were classified according to the events of the news by a term classifier. The main function of the automatic fuzzy inference mechanism was to generate the membership degrees (classification) for each event with respect to the fuzzy concepts defined in the fuzzy ontology. The standard triangular membership function was used for the classification purpose. The method discussed in this paper is a fully automatic fuzzy domain ontology discovery approach. There is no predefined fuzzy concepts and taxonomy of concepts, instead our fuzzy domain ontology extraction method will automatically discover the concepts and generate the taxonomy relations. In addition, there is no need to set the artificial threshold values for the triangular membership function, instead our membership function can automatically derive the membership values based on the lexico-syntactic and statistical features of the terms observed in a textual database.

An ontology mining technique was proposed to extract patterns representing users' information needs [29]. The ontology mining method consists of two parts: the top backbone and the base backbone. The former represents the relations between compound classes of the ontology. The latter indicates the linkage between primitive classes and compound classes. The Dempster-Shafer theory of evidence model was applied to extract the relations among classes. The strength of the ontology mining method is that it can effectively synthesize taxonomic relations and non-taxonomic relation in a single ontology model. In addition, a novel method was proposed to capture the evolving patterns in order to refine the initially discovered ontology. Finally, a formal model was developed to assess the relevance of the discovered ontology with respect to the user's information needs. The ontology mining method was validated based on the Reuters RCV-1 benchmark collection. The research work presented in this paper focuses on fuzzy domain ontology discovery rather than the discovery of crisp ontology representing users' information needs. Instead of using Dempster-Shafer theory of evidence, our concept extraction method is underpinned by information theoretic approaches such as mutual information.

Sanderson and Croft [45] proposed a document-based subsumption induction method to automatically derive a hierarchy of terms from a corpus. In particular, the subsumption relations among terms are developed based on the co-occurrence of terms in the documents of a corpus. For example, term $t_1$ is considered more specific than another term $t_2$ if the appearance of $t_1$ in a document implies the appearance of $t_2$ in the same document but not vice versa. They adopted an artificial threshold such as $Pr(t_2|t_1) \geq 0.8$ as a fixed cut-off to determine the specificity relation between $t_1$ and $t_2$. Even though the idea is interesting, the computational method may not be robust enough to deal with taxonomy extraction tasks in general. Our method differs from their work in that we are dealing with the more challenging task of concept hierarchy extraction rather than term relationship extraction. In addition, our method extends their computational method in that the co-occurrence of terms is derived based on a moving text window rather than the whole document to reduce the chance of generating noisy subsumption relations. Our method is more robust than their approach because there is no need of specifying an artificial threshold to establish concept specificity relation.

An ontology discovery approach is proposed to improve domain ontologies by mining the hidden semantics from text [10]. The learning approach is based on self-organizing map (SOM). The words occurring in free-form text documents from the application domain are clustered according to their semantic similarity based on statistical context analysis. A word is described by words that appear within a fix-sized context window, semantic relations of words are then extracted and represented in the self-organizing map. As a result, words that refer to similar objects are found in neighboring parts of the map. The two dimensional map representation provides an intuitive interface for browsing through the vocabulary to discover new concepts or relations between concepts that are still missing in the ontology. It is argued that such an approach is suitable for finding new concepts and relations to be added to the associative network. The SOM approach was illustrated with reference to the tourism domain and a field test based on the largest Austrian tourism Web site was conducted to validate the ideas. Our ontology extraction method is based on context-sensitive text mining and fuzzy relation construction rather than using SOM.

## III. A Framework for Automatic Concept Map Generation

It has been pointed out that the main challenge of automatic ontology extraction from textual databases is the removal of noisy concepts and relations[30], [31]. Based on this premise, our domain ontology extraction methodology in general and the concept map generation process in particular are designed to effectively filter the non-relevant concepts and concept relations from the concept space. Figure 2 depicts the proposed methodology of automatically generating concept maps from a collection of online messages posted to blogs, emails, chat rooms, Web pages, etc. The collection of messages is treated as a textual corpus. At the document parsing stage, our document parser will scan each message to analyze the lexico-syntactic elements embedded in the message. For instance, stop words such as "a, an, the" are removed from the message since these words appear in any contexts and they cannot provide useful information to describe a domain concept. For our implementation, a stop word file is constructed based on the standard stop word file used in the SMART retrieval system [43]. Different customizations is required for processing different kinds of documents. For example, we need to extend the SMART stop word file by including stop words such as "home", "contact", "web", "site", etc. for parsing Web pages.

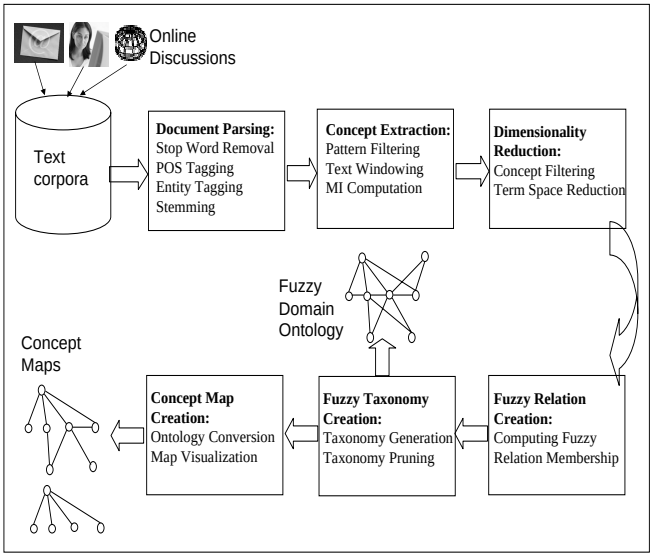


Fig. 2. A Framework for Automatic Concept Map Generation

Lexical patterns are identified by applying Part-of-Speech (POS) tagging to the source documents. We develop our POS tagger based on the WordNet lexicon and the publicly available API (`http://wordnet.princeton.edu/`). For named-entity detection (e.g., people's names, organizations' names, etc.), we employ BBN's IdentiFinder [1]. However, for the e-Learning application reported in this paper, we do not make use of the entity tags for concept extraction. We simply treat each named-entity as a noun for subsequent linguistic pattern mining. After the tagging process, each token is stemmed according to the Porter stemming algorithm [40]. During the concept extraction stage (Section V-A), certain linguistic patterns are ignored to reduce the generation of noisy concepts. For example, ontology engineers or instructors in the case of e-Learning application, will specify the mining focus on certain linguistic patterns such as "Noun Noun", "Adjective Noun", "Verb Noun", etc. The text mining program will then focusing on finding the term association information and collecting the statistical data for those patterns only. Not only does it reduce the generation of noisy concepts but also improve the computational efficiency of our ontology extraction process. A text windowing process will be conducted by scanning adjacent tokens within a pre-defined window size of 5 to 10 words from left to right over all the documents. At the end of the windowing process, an information theoretic measure

is applied to compute the co-occurrence statistics between the targeting linguistic patterns and other tokens appearing in the same text window across the corpus. Thereby, context vectors can be created to describe the semantic of the extracted concepts.

In addition, to filter out non-relevant domain concepts, the occurrence of a concept across different domains (e.g., corpora) will be assessed (Section V-B). The basic intuition is that a concept frequently appears in a specific domain (corpus) rather than many different domains is more likely to be a relevant domain concept. Those concepts with relevance scores below a certain threshold will not be used for taxonomy generation. To produce accurate concept representations, a dimensionality reduction method is applied to the filtered concept space to minimize the terms (features) used to characterize the concepts based on the principle of minimal information loss (Section V-C). After concept space reduction, the subsumption relationships among the domain concepts are computed according to our fuzzy relation membership function (Section V-D). A taxonomy of fuzzy domain concepts is then constructed according to our fuzzy domain ontology extraction algorithm (Section V-E). Finally, our visualization mechanism converts the fuzzy domain ontology to concept maps and displays them on our Web-based e-Learning platform.

Before illustrating the computational details of our fuzzy domain ontology extraction method in the remaining sections, we should give a precise definition of what we mean light weight fuzzy domain ontology In particular, our proposed model of fuzzy domain ontology is underpinned by fuzzy sets and fuzzy relations [60].

*Definition 1 (Fuzzy Set):* A fuzzy set $\mathcal{F}$ consists of a set of objects drawn from a domain $X$ and the membership of each object $x_i$ in $\mathcal{F}$ is defined by a membership function $\mu_{\mathcal{F}} : X \mapsto [0, 1]$. If $Y$ is a crisp set, $\varphi(Y)$ denotes a fuzzy set generated from the traditional set of items $Y$.

*Definition 2 (Fuzzy Relation):* A fuzzy relation $R_{XY}$ is defined as the fuzzy set $\mathcal{R}$ on a domain $X \times Y$ where $X$ and $Y$ are two crisp sets. The membership of each object $(x_i, y_i)$ in $\mathcal{R}$ is defined by a membership function $\mu_{\mathcal{R}} : X \times Y \mapsto [0, 1]$.

*Definition 3 (Fuzzy Ontology):* A fuzzy ontology is a 6-tuple $Ont = \langle X, A, C, R_{XC}, R_{AC}R_{CC}\rangle$, where X is a set of objects, A is the set of attributes describing the objects, and C is a set of concepts (classes). The fuzzy relation $R_{XC} : X \times C \mapsto [0, 1]$ assigns a membership to the pair $(x_i, c_i)$ for all $x_i \in X, c_i \in C$, the fuzzy relation $R_{AC} : A \times C \mapsto [0, 1]$ define the mapping from the set of attributes $A$ to the set of concepts $C$, and the fuzzy relation $R_{CC} : C \times C \mapsto [0, 1]$ define the strength of the sub-class/super-class relationships among the set of concepts $C$.

Figure 3 illustrates our formal model of light weight fuzzy domain ontology with reference to the above definitions. In this example, $X = \{x_1, \ldots, x_7\}, A = \{a_1, \ldots, a_6\}$, and $C = \{c_1, \ldots, c_5\}$ are assumed. The fuzzy relations among concepts (i.e., sub-class/super-class relationships) is denoted

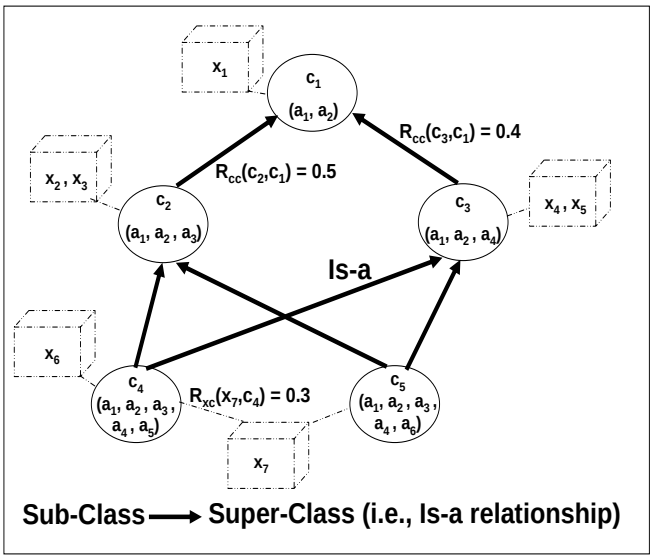


Fig. 3. A Formal Model of Fuzzy Domain Ontology

$R_{CC}(c_x, c_y)$ and two examples e.g., $R_{CC}(c_2, c_1) = 0.5$ and $R_{CC}(c_3, c_1) = 0.4$ are shown in Figure 3. The $R_{XC}$ relation describes the membership of an object for a particular class (concept). For instance, object $x_7$ is considered belonging to the class $c_4$ with a membership value of $0.3$. For the e-Learning application, the ontology can represent which online message (i.e., an object) created by a learner is associated with certain concepts to facilitate the analysis of students' understanding. To improve the readability of Figure 3, the partial associations between concepts and attributes (i.e. $R_{AC}$) are not depicted. For a concept such as "commercial bank", we may find a property term (i.e., attribute) "customer" describing the concept. However, the term "customer" may also be used to describe other concepts such as "book shop" to a certain degree. Our light weight fuzzy domain ontology model is able to represent the partial association between concepts and the underlying property terms. Based on the idea of formal concept analysis [8], $X$ is the *extent* of the concepts $C$, and $A$ is the *intent* which defines the properties of $C$. According to the concept of subsumption, the sub-concept/super-concept relation ($R_{CC}$) can be defined by:

*Definition 4 (Fuzzy Subsumption):* With respect to an arbitrary $\alpha$-cut level, a concept $c_x \in C$ is the sub-concept of another super-concept $c_y \in C$ if and only if $\forall a_i \in \{z \in A | \mu_{R_{AC}}(z, c_y) \geq \alpha\}$, $\mu_{R_{AC}}(a_i, c_x) \geq \alpha$. Alternatively, from an extensional perspective, a concept $c_x \in C$ is the sub-concept of another super-concept $c_y \in C$ if and only if $\forall x_i \in \{z \in X | \mu_{R_{XC}}(z, c_x) \geq \alpha\}$, $\mu_{R_{XC}}(x_i, c_y) \geq \alpha$ with respect to an arbitrary $\alpha$-cut level.

Definition 4 can be explained as follows: if the membership of every attribute $a_i \in A$ for the concept $c_y \in C$ is greater than or equal to a certain threshold $\alpha$, the membership of the corresponding attribute $a_i$ for the concept $c_x \in C$ is also

greater than or equal to $\alpha$, then the concept $c_x$ is the sub-concept of $c_y$. As can be seen, the crisp subsumption relation is only a special case of the generalized fuzzy subsumption relation in that the threshold value $\alpha = 1$ is established for the special case. In other words, if it is true that every attribute $a_i \in A$ characterizing the concept $c_y$ implies that it also characterizes the concept $c_x$, the concept $c_x$ is the sub-concept of $c_y$.

## IV. The Linguistic Foundations

From a human cognitive perspective, humans acquire the meaning of a new concept by associating the contexts in which the concept appears. Our concept extraction method is developed based on such an intuition. In particular, our context-sensitive text mining approach is based on the *distributional hypothesis* which assumes that terms (concepts) are similar according to the extent that they share similar linguistic contexts [15]. In particular, we borrow the notion of *collocational expressions* from computational linguistic to extract the semantics of certain lexical elements (e.g., concepts) from text corpora. In computational linguistic, a term refers to one or more tokens (words), and a term could also been seen as a concept if it carries recognizable meaning with respect to a context (domain) [17], [35]. Collocational expressions are groups of words related in meaning, and the constituent words of an expression are frequently found in a near loci of a few adjacent words in a textual unit [47], [50]. Collocational expressions provide the contexts to extract the semantics of concepts embedded in natural language texts such as net news, blogs, emails, or Web documents.

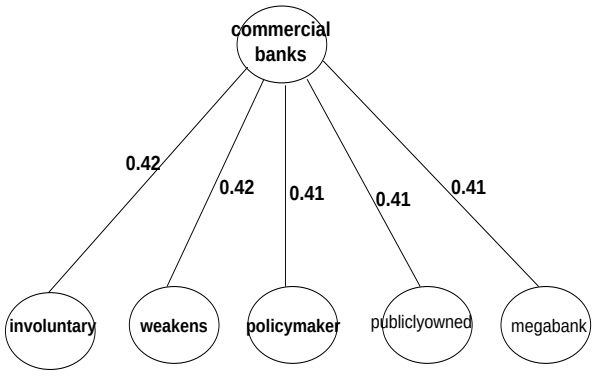


Fig. 4. Domain Specific Semantics of the Concept "Commercial Banks"

In the field of information retrieval (IR), the notion of *context vectors* [17], [46] has been proposed to construct computer-based representations of concepts (i.e., linguistic class). In this approach, a concept is represented by a vector of words (property terms) and their numerical weights. The weight of a word indicates the extent to which the particular word is *associated* with the underlying concept. Figure 4 shows that the concept "commercial banks" is represented by the property terms such as involuntary, weakens, policymaker, publiclyowned, megabank, etc. Indeed, this is an interesting example generated from the Reuters-21578 corpus (http://www.daviddlewis.com/resources/testcollections/). The context vector of "commercial banks" is shown as follows:

Concept: commercial banks
Context Vector:
$\langle (involuntary, 0.42), (weakens, 0.42), (policymaker, 0.41), (publiclyowned, 0.41), (megabank, 0.41) \rangle$

A context vector can be seen as a point in a multi-dimensional geometric information space with each dimension representing a property term. It should be noted that the meanings (senses) of "comercial banks" is "a financial institution that accepts deposits and makes loans and provides other services for the public" as defined in WordNet [33], which is quite different from that discovered by our context-sensitive text mining method [24]. Static lexicons such as WordNet can only capture the lexical knowledge of a concept, but fails to represent context-sensitive information relevant for a specific domain. In this example, the Reuters-21578 corpus describes the domain of the U.S. financial market in 1987. A linguistic concept such as "commercial banks" can be taken as a class (set) with respect to the fuzzy sets framework. A property term such as "publiclyowned" can then be treated as an attribute describing the concept to a certain degree (i.e., $\mu_{R_{AC}}(publiclyowned, commercialbanks) = 0.41$) .

## V. Automatic Fuzzy Domain Ontology Extraction

### A. Concept Extraction

Our text mining method is specifically designed to filter noisy concepts. After standard document pre-processing such as stop word removal, POS tagging, and word stemming [44], a *windowing process* is conducted over the collection of documents. The windowing process can help reduce the number of noisy terms. For each document (e.g., Net news, Web page, email, etc.), a *virtual window* of $\delta$ words is moved from left to right one word at a time until the end of a textual unit (e.g., a sentence) is reached. Within each window, the statistical information among tokens is collected to develop collocational expressions. Such a windowing process has successfully been applied to text mining before [24]. The windowing process is repeated for each document until the entire collection has been processed. According to previous studies, a text window of 5 to 10 terms is effective [17], [39], and so we adopt this range as the basis to perform our windowing process. To improve computational efficiency and filter noisy concepts, only the specific linguistic patterns (e.g., Noun Noun, Adjective Noun, etc.) defined by the user will be analyzed. After parsing the whole corpus, the statistical data (e.g., mutual information) about the potential concepts is collected by our statistical token analyzer. If the association weight between a concept and a term is below a pre-defined threshold value $\zeta$, it will be discarded from the context vector of the concept.

For statistical token analysis, several information theoretic methods are employed. Mutual Information has been applied to collocational analysis [39], [52] in previous research. Mutual Information is an information theoretic method to compute the dependency between two entities and is defined by [48]:

$$MI(t_i, t_j) = \log_2 \frac{Pr(t_i, t_j)}{Pr(t_i)Pr(t_j)} \tag{1}$$

where $MI(t_i, t_j)$ is the mutual information between term $t_i$ and term $t_j$. $Pr(t_i, t_j)$ is the joint probability that both terms appear in a text window, and $Pr(t_i)$ is the probability that a term $t_i$ appears in a text window. The probability $Pr(t_i)$ is estimated based on $\frac{|w_t|}{|w|}$ where $|w_t|$ is the number of windows containing the term $t$ and $|w|$ is the total number of windows constructed from a corpus. Similarly, $Pr(t_i, t_j)$ is the fraction of the number of windows containing both terms out of the total number of windows.

We develop *Balanced Mutual Information* (BMI) to compute the degree of association among tokens. This method considers both term presence and term absence as the evidence of the implicit term relationships.

$$\begin{aligned}\mu_{c_i}(t_j) &\approx BMI(t_i, t_j) \\ &= \beta \times [Pr(t_i, t_j)\log_2(\tfrac{Pr(t_i,t_j)+1}{Pr(t_i)Pr(t_j)}) + \\ &\quad Pr(\neg t_i, \neg t_j)\log_2(\tfrac{Pr(\neg t_i,\neg t_j)+1}{Pr(\neg t_i)Pr(\neg t_j)})] - \\ &\quad (1-\beta) \times [Pr(t_i, \neg t_j)\log_2(\tfrac{Pr(t_i,\neg t_j)+1}{Pr(t_i)Pr(\neg t_j)}) + \\ &\quad Pr(\neg t_i, t_j)\log_2(\tfrac{Pr(\neg t_i,t_j)+1}{Pr(\neg t_i)Pr(t_j)})]\end{aligned} \tag{2}$$

where $\mu_{c_i}(t_j)$ is the membership function to estimate the degree of a term $t_j \in A$ belonging to a concept $c_i \in C$. $\mu_{c_i}(t_j)$ is the computational mechanism for the relation $R_{AC}$ defined in the fuzzy domain ontology $Ont = \langle X, A, C, R_{XC}, R_{AC}R_{CC}\rangle$. The membership function $\mu_{c_i}(t_j)$ is indeed approximated by the BMI score. The weight factor $\beta > 0.5$ is used to control the relative importance of two kinds of evidence (positive and negative).

Other measures that are used to estimate the membership values of $t_j \in c_i$ include Jaccard (JA), conditional probability (CP), Kullback-Leibler divergence (KL), and Expected Cross Entropy (ECH) [21], and Normalized Google Distance (NGD) [7]. For Eq.(7), the term $|w_{c_i}|$ means the number virtual text windows containing the concept $c_i$ and the term $|w_{c_i,t_j}|$ refers to the number of virtual text windows containing both $c_i$ and $t_j$. After computing term-concept association weights using any one of the methods mentioned above, the association weights are subject to linear scaling using $v_{Norm} = \frac{v - v_{min}}{v_{max} - v_{min}}$. As a result, all the term-concept association weights fall into the unit interval $\forall_{c_i \in C, t_j \in X} \mu_{c_i}(t_j) \in [0, 1]$. As NGD is a distance measure, we would use the dual function to generate the membership values (e.g., $\mu_{c_i}(t_j) = 1 - NGD_{Norm}(c_i, (t_j))$. A $\zeta$-cut is applied to discard terms from the potential concept if their memberships are below the threshold $\zeta$. It should be noted that the constituent terms of a concept are always implicitly associated with the concept itself with the maximal membership 1.

$$\begin{aligned}\mu_{c_i}(t_j) &\approx Jacc(c_i, t_j) \\ &= \tfrac{Pr(c_i \wedge t_j)}{Pr(c_i \vee t_j)}\end{aligned} \tag{3}$$

$$\begin{aligned}\mu_{c_i}(t_j) &\approx Pr(c_i|t_j) \\ &= \tfrac{Pr(c_i, t_j)}{Pr(t_j)}\end{aligned} \tag{4}$$

$$\begin{aligned}\mu_{c_i}(t_j) &\approx KL(c_i||t_j) \\ &= \textstyle\sum_{c_i \in C} Pr(c_i|t_j)\log_2 \tfrac{Pr(c_i|t_j)}{Pr(c_i)}\end{aligned} \tag{5}$$

$$\begin{aligned}\mu_{c_i}(t_j) &\approx ECH(t_j, c_i) \\ &= Pr(t_j)\textstyle\sum_{c_i \in C} Pr(c_i|t_j)\log_2 \tfrac{Pr(c_i|t_j)}{Pr(c_i)}\end{aligned} \tag{6}$$

$$\begin{aligned}\mu_{c_i}(t_j) &\approx NGD(c_i, t_j) \\ &= \tfrac{\max\{\log_2|w_{c_i}|, \log_2|w_{t_j}|\} - \log_2|w_{c_i,t_j}|}{\log_2|w+1| - \min\{\log_2|w_{c_i}|, \log_2|w_{t_j}|\}}\end{aligned} \tag{7}$$

### *B. Concept Filtering*

To further filter the noisy concepts, we adopt the TFIDF [44] like heuristic to perform the filtering process. Similar approach has also been used in ontology learning [36]. For example, if a concept is significant for a particular domain, it will appear more frequently in that domain when compared with its appearance in other domains. The following measure is used to compute the relevance score of a concept:

$$Rel(c_i, D_j) = \frac{Dom(c_i, D_j)}{\sum_{k=1}^{n} Dom(c_i, D_k)} \tag{8}$$

where $Rel(c_i, D_j)$ is the relevance score of a concept $c_i$ in the domain $D_j$. The term $Dom(c_i, D_j)$ is the domain frequency of the concept $c_i$ (i.e., number of documents containing the concept divided by the total number of documents in the corpus). The higher the value of $Rel(c_i, D_j)$, the more relevant the concept is for domain $D_j$. Based on empirical testing, we can estimate a threshold $\varpi$ for a particular domain. Only the concepts with relevance scores greater than the threshold will be selected. For each selected concept, its context vector will be expanded based on the synonymy relation defined in WordNet [33]. This is in fact a *smoothing* procedure [8]. The intuition is that some terms characterizing a particular concept may not co-occur with the concept in a corpus. To make our ontology extraction method more robust, we need to consider these missing properties. For instance, the context vector "commercial banks" of our example will be expanded with the term "deposits" based on the synonymy relation of WordNet, and a default membership will be assigned to such a term.

### *C. Dimensionality Reduction*

In order to reduce the terms dimensionality, unsupervised mapping techniques to lower dimension, for example, *Principal Component Analysis* (PCA) [18] and *Singular Value Decomposition* (SVD) [12] [11] can be applied to the *Term-Concept Association Matrix*, $\mathbf{R}$, which is formed by the membership values $\mu_{c_i}(t_j)$ for all term $t_j \in A$ belonging to some concepts $c_i \in C$ after the previous empircal process of noisy and irrelevant concepts reduction. To alleviate the burden

in computing the covariance matrix in PCA, we decompose the Term-Concept matrix $\mathbf{R}$ using SVD. $\mathbf{R}$ can be expressed as any rectangular $m \times n$ matrix. The general complexity of computing SVD is in $O(min(mn^2, nm^2)$. As the number of concepts has been reduced to k (where $k << n$) by the concept filtering process, the actual computational complexity of our SVD process is reduced to $O(min(mk^2, km^2)$. In addition, by controlling the filtering parameter $\varpi$, our SVD can scale up for a large collection of messages. Element $(j, i)$ in $\mathbf{R}$ represents the membership value $\mu_{c_i}(t_j)$, i.e.

$$\mathbf{R} = \begin{pmatrix} \mu_{c_1}(t_1) & \mu_{c_2}(t_1) & \dots & \mu_{c_i}(t_1) & \dots & \mu_{c_n}(t_1) \\ \mu_{c_1}(t_2) & \mu_{c_2}(t_2) & \dots & \mu_{c_i}(t_2) & \dots & \mu_{c_n}(t_2) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mu_{c_1}(t_j) & \mu_{c_2}(t_j) & \dots & \mu_{c_i}(t_j) & \dots & \mu_{c_n}(t_j) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mu_{c_1}(t_m) & \mu_{c_2}(t_m) & \dots & \mu_{c_i}(t_m) & \dots & \mu_{c_n}(t_m) \end{pmatrix} \tag{9}$$

Each row $\mathbf{R_j}$ in the association matrix $\mathbf{R}$ is a vector corresponding to the membership degree of a term $t_j$ belonging to each concept $c_i$ of the reduced concept space $C$, $\forall c_i \in C|_{i=1,\dots,n}$.

$$\mathbf{R_j^T} = \begin{pmatrix} \mu_{c_1}(t_j) & \mu_{c_2}(t_j) & \dots & \mu_{c_i}(t_j) & \dots & \mu_{c_n}(t_j) \end{pmatrix} \tag{10}$$

Similarly, each column $\mathbf{R_i}$ in $\mathbf{R}$ is a vector corresponding to a concept $c_i$ giving various degrees of each term $t_j$, $\forall t_j \in A|_{j=1,\dots,m}$.

$$\mathbf{R_i} = \begin{pmatrix} \mu_{c_i}(t_1) & \mu_{c_i}(t_2) & \dots & \mu_{c_i}(t_j) & \dots & \mu_{c_i}(t_m) \end{pmatrix} \tag{11}$$

By SVD, $\mathbf{R}$ can be decomposed into the product of three other matrices:

$$\mathbf{R} = \mathbf{USV^T} \tag{12}$$

where $\mathbf{S}$ is a $l \times l$ diagonal matrix such that $\mathbf{S} = [\delta_{i,j}]$, where $\forall_{i=j}\ \delta_{i,j} \neq 0$ and $\forall_{i \neq j}\ \delta_{i,j} = 0$, $\mathbf{U}$ and $\mathbf{V}$ have orthogonal and unitary columns such that $\mathbf{U^T U} = \mathbf{I}, \mathbf{V^T V} = \mathbf{I}$, $\mathbf{I}$ is the identity matrix.

$\mathbf{R}$ can be expressed as:

$$\mathbf{R} = \begin{pmatrix} [\mathbf{u}_1] & \dots & [\mathbf{u}_l] \end{pmatrix} \cdot \begin{pmatrix} \delta_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \delta_l \end{pmatrix} \cdot \begin{pmatrix} [\mathbf{v}_1] \\ \vdots \\ [\mathbf{v}_l] \end{pmatrix} \tag{13}$$

$\mathbf{U} = ([\mathbf{u_j}])|_{j=1,\dots,l}$ and $\mathbf{V} = ([\mathbf{v_i}])|_{i=1,\dots,l}$ are the left and right singular vectors, respectively, corresponding to the monotonically decreasing singular values $\delta_l | l \in (i, j), i = j$ of the diagonal matrix $\mathbf{S}$ and $l = \min(m, n)$. The full-rank or $l$-rank $\mathbf{R}$ could then be approximated to a rank-$k$ approximation using Latent Semantics Structure [12] in which the largest $k$ singular values of $\mathbf{R}$ associated with the first $k$ columns of the $\mathbf{U}$ and $\mathbf{V}$ matrices are used for reconstruction, i.e. $\mathbf{R}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T$ where $k << l$. Therefore, the $\mathbf{R}_k$ is the closest rank-$k$ approximation in term of least square error sense as it consists of the $k$ largest singular triplets of $\mathbf{R}$ [57]. As a result, a new set of membership value $\mu_{c_i}(t_j)$ will then represent the degree of a term $t_j$ in the reduced term space with only $k$-dimensions, for the concepts $c_i \in \{c_1, \dots, c_n\}$.

### *D. Fuzzy Relation Extraction*

The final stage towards our ontology extraction method is fuzzy taxonomy generation based on the subsumption relations among the extracted concepts. Let $Spec(c_x, c_y)$ denotes that concept $c_x$ is a specialization (sub-class) of another concept $c_y$. The degree of such a specialization relation can be estimated from:

$$\begin{aligned} \mu_{R_{CC}}(c_x, c_y) &\approx Spec(c_x, c_y) \\ &= \frac{\sum_{t \in c_x \cap c_y} \mu_{c_x}(t) \otimes \mu_{c_y}(t)}{\sum_{t_x \in c_x} \mu_{c_x}(t_x)} \end{aligned} \tag{14}$$

where $\otimes$ is a fuzzy conjunction operator which is equivalent to the min function. The above formula states that the degree of subsumption (specificity) of $c_x$ to $c_y$ is based on the ratio of the sum of the minimal membership values of the common terms belonging to both concepts to the sum of the membership values of terms in the concept $c_x$. For instance, if every attribute of $c_y$ is also an attribute of $c_x$, a strong specificity relation exists and the value of $Spec(c_x, c_y)$ is high. The domain of the $Spec(c_x, c_y)$ falls in the unit interval $[0, 1]$ and the subsumption relation is asymmetric. Eq.(14) has been applied to our earlier studies of fuzzy ontology extraction [23], [26], [25].

One problem of the standard fuzzy conjunction operation is that the specificity value is highly influenced by the weakest terms (attributes) of the concepts. Therefore, we explore another alternative of estimating the degree of subsumption between two concepts based on the method successfully applied to image analysis [55]. In particular, any two concepts $c_x$ and $c_y$ could be said to be similar if their structural similarity is high and the corresponding structural similarity value $SSIM(c_x, c_y)$ approaches 1 [55]. On the other hand, two concepts are dissimilar if their structural similarity value $SSIM(c_x, c_y)$ is low (e.g., close to zero). The $SSIM(c_x, c_y)$ function is expressed as:

$$SSIM(c_x, c_y) = l(c_x, c_y) \cdot c(c_x, c_y) \cdot s(c_x, c_y) \tag{15}$$

$$l(c_x, c_y) = \frac{2M_{c_x}M_{c_y} + Q_1}{M_{c_x}^2 + M_{c_y}^2 + Q_1} \tag{16}$$

$$c(c_x, c_y) = \frac{2\sigma_{c_x}\sigma_{c_y} + Q_2}{\sigma_{c_x}^2 + \sigma_{c_y}^2 + Q_2} \tag{17}$$

$$s(c_x, c_y) = \frac{\sigma_{c_x, c_y} + Q_3}{\sigma_{c_x}\sigma_{c_y} + Q_3} \tag{18}$$

where $c_x, c_y \in C$. For our application, the $l(c_x, c_y)$ function is used to measure the similarity of two concepts in terms of *semantic coherence*, whereas the $c(c_x, c_y)$ function is used to estimate the similarity between two concepts in terms of *semantic variance*. Finally, the $s(c_x, c_y)$ function is applied to measure the similarity of two concepts based

on their component structures. Slightly different from [55], our similarity metric is applied to the concept vectors $c_i = \langle \mu_{c_i}(t_1), \ldots, \mu_{c_i}(t_m) \rangle$. The mean and standard deviation of each concept vector, and the covariance between two concept vectors are defined by:

$$M_{c_i} = \frac{1}{m}(\sum_{j=1}^{m} \mu_{c_i}(t_j)) \tag{19}$$

$$\sigma_{c_i} = (\frac{1}{m-1}(\sum_{j=1}^{m}(\mu_{c_i}(t_j) - M_{c_i})^2))^{1/2} \tag{20}$$

$$\sigma_{c_x,c_y} = (\frac{1}{m-1}(\sum_{j=1}^{m}(\mu_{c_x}(t_j) - M_{c_x})(\mu_{c_y}(t_j) - M_{c_y}))) \tag{21}$$

The terms $Q_1 = 0.0255$, $Q_2 = 0.2295$ and $Q_3 = 0.1148$ are constants and they are applied to image analysis work before [55]. We adopt $Q_1 = [0,5 \times 0.0255], Q_2 = [0,5 \times 0.2295], Q_3 = [0,5 \times 0.1148]$ in our experiments. When we apply the structural similarity measure to estimate the degree of subsumption between two concepts, we follow the same intuition illustrated in Definition 4. For instance, if most attributes $t_i$ belonging to the concept $c_y$ are also belonging to the concept $c_x$, the concept $c_x$ is a sub-concept of $c_y$ to a high degree. To formulate our $Spec(c_x, c_y)$ function based on the structural similarity, we first compute the common concept $c_g = c_x \cap c_y$. Then, we examine if this common sub-concept is more subsumed by which concept to determine the direction of the specialization relation. Thereby, the degree of specificity from $c_x$ to $c_y$ is approximated by:

$$\begin{aligned} \mu_{R_{CC}}(c_x, c_y) &\approx Spec(c_x, c_y) \\ &= \begin{cases} 0 \quad \text{if } SSIM(c_y, c_g) > SSIM(c_x, c_g) \\ \frac{SSIM(c_x,c_g) - SSIM(c_y,c_g)}{SSIM(c_x,c_g)} \quad \text{otherwise} \end{cases} \end{aligned} \tag{22}$$

The above formula states that the degree of subsumption (specificity) of $c_x$ to $c_y$ is based on the ratio of the difference between the structural similarity of $SSIM(c_x, c_g)$ and $SSIM(c_y, c_g)$ to the normalization factor $SSIM(c_x, c_g)$. On the other hand, if more common structural elements are found in $c_y$ rather than $c_x$ (i.e., $c_y$ is a sub-concept of $c_x$), the degree of the specificity relation from $c_x$ to $c_y$ is zero.

### *E. Fuzzy Taxonomy Extraction*

When the taxonomy is built, we only select the subsumption relations such that $Spec(c_x, c_y) \geq Spec(c_y, c_x)$ and $Spec(c_x, c_y) > \lambda$ where $\lambda$ is a threshold to distinguish significant subsumption relations. The parameter $\lambda$ is estimated based on empirical tests. If $Spec(c_x, c_y) = Spec(c_y, c_x)$ and $Spec(c_x, c_y) > \lambda$ is established, the *equivalent* relation between $c_x$ and $c_y$ will be extracted. In addition, a pruning step is introduced such that the redundant taxonomy relations are removed. If the membership of a relation $\mu_{C \times C}(c_1, c_2) \leq \min(\{\mu_{C \times C}(c_1, c_i), \ldots, \mu_{C \times C}(c_i, c_2)\})$, where $c_1, c_i, \ldots, c_2$ form a path $P$ from $c_1$ to $c_2$, the relation $R_(c_1, c_2)$ is removed because it can be derived from other stronger taxonomy relations in the ontology. The fuzzy domain ontology mining algorithm is summarized and shown in Figure 5. According to this algorithm, more than one connected graph could be generated from a corpus. Each graph will be used as the basis to generate a concept map. The general computational complexity of our algorithm is characterized by $O(k^2 m + km^2)$, where $k$ is the reduced dimensionality of the term space, and $m$ is the reduce cardinality of the set $C$. By controlling the concept filtering threshold $\varpi$, we can turn $m$ into a small number. Moreover, we can make a trade-off between computational time and accuracy by tuning $k$ during dimensionality reduction. As a result, our algorithm can scale up even for a large number of messages. We have conducted a field test to demonstrate that our system can run efficiently in practice.

**Algorithm FuzzyOntoExt(*D, PA, Ont*)**
**Input:** corpus $D$ and vector of threshold values $PA$
**Output:** a light weight fuzzy domain ontology $Ont$
**Main Procedure:**

1) $Ont = \{\}$
2) For each document $d \in D$ Do
   a) Construct text windows $w \in d$
   b) Remove stop words $sw$ from $w$
   c) Perform POS tagging for each term $t_i \in w$
   d) Apply Porter stemming to each term $t_i$
   e) Filter specific linguistic patterns
   f) Accumulate the frequency for $t_i \in w$ and the joint frequency for any pair $t_i, t_j \in w$
   g) IF $lower \leq Feq(t_i) \leq upper$, $A = A \cup t_i$
3) For each term $t_i \in A$ Do
   a) compute its context vector $c_i$ using BMI, MI, JA, CP, KL, ECH, or NGD
   b) $C = C \cup c_i$
4) For each $c_i \in C$ Do /* Concept Pruning - $\alpha$-cut */
   a) IF $\exists t_i \in c_i : \mu_{c_i}(t_i) < \zeta$
   b) THEN $C = C - c_i$
5) $\forall c_i \in C$ : Compute $Rel(c_i, D_j)$
6) IF $Rel(c_i, D_j) < \varpi$ /* Concept Filtering */
7) THEN $C = C - c_i$
8) Perform Dimensionality Reduction SVD
9) For each pair of concepts $c_i, c_j \in C$ Do
   a) Compute the taxonomy relation $r(c_i, c_j)$ using $Spec(c_i, c_j)$
   b) IF $\mu_{R_{CC}}(c_i, c_j) > \lambda$, $R_{CC} = R_{CC} \cup r(c_i, c_j)$
10) For each $r(c_i, c_j) \in R$ Do /* Taxonomy Pruning */
   a) IF $\mu_{R_{CC}}(c_i, c_j) < \mu_{R_{CC}}(c_j, c_i)$
   b) THEN $R_{CC} = R_{CC} - r(c_i, c_j)$
   c) IF $\exists P(c_i \rightarrow c_x, \ldots, c_y \rightarrow c_j)$
   d) AND $\mu_{R_{CC}}(c_i, c_j) \leq \min(\{\mu_{R_{CC}}(c_i, c_x), \mu_{R_{CC}}(c_x, c_y), \ldots, \mu_{R_{CC}}(c_y, c_j)\})$
   e) THEN $R_{CC} = R_{CC} - r(c_i, c_j)$
11) Output $Ont$

Fig. 5. The Fuzzy Domain Ontology Extraction Algorithm

## VI. Application to e-Learning

In an e-Learning environment, learners are often encouraged to reflect what they have learned by writing online journals or sharing their ideas via an online discussion board. Figure 6 shows a sample of message entered by a student via

the Blackboard e-learning environment. Usually, instructor or other fellow students may reply and produce multiple threads of messages like the one shown in Figure 7. If an instructor wants to know the current learning status of their students, she need to browse through all the threads of messages to analyze the contents slowly. Given the fact that humans' cognitive power is quite limited, such a mental analysis process is very time consuming, and it is very unlikely that the instructor can do it on the fly (i.e., when a lecture or tutorial is in progress). To alleviate such a problem and to facilitate adaptive teaching and learning, we can apply the fuzzy domain ontology discovery algorithm to automatically extract and visualize the concept maps representing an individual or a group of learners' knowledge structure. Based on the concept maps, the instructor can examine whether the existing concepts have been thoroughly internalized by her students or not, and then she can decide which topics should be covered next.

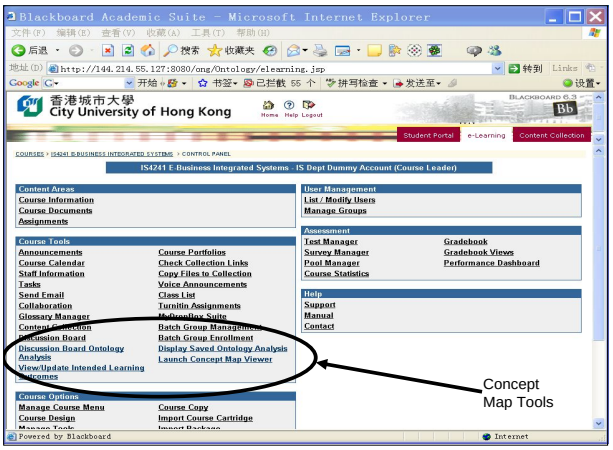


Fig. 8. The Concept Map Generation Tool on an e-Learning Platform

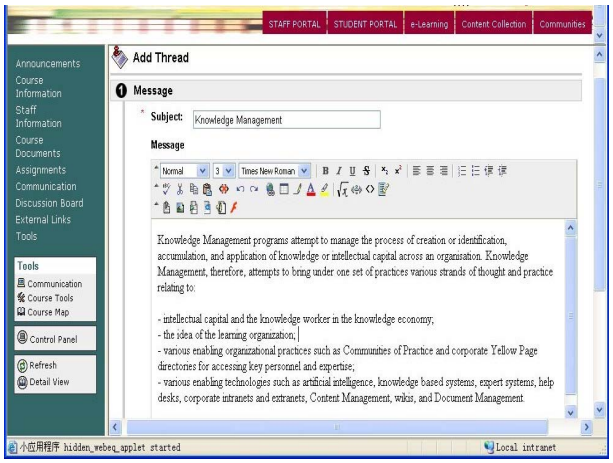


Fig. 6. A Message Entered via Online Discussion Board

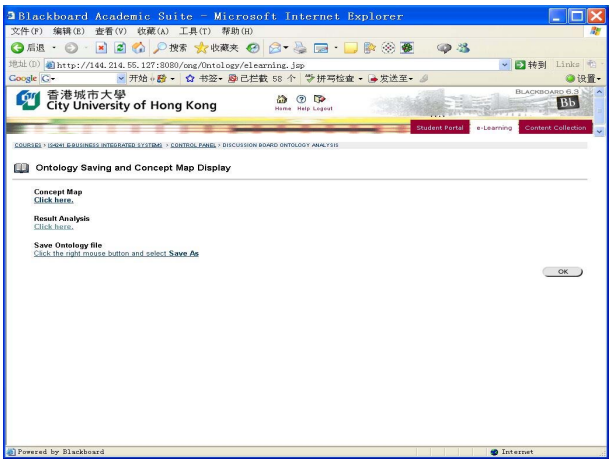


Fig. 9. The Display Concept Map Panel

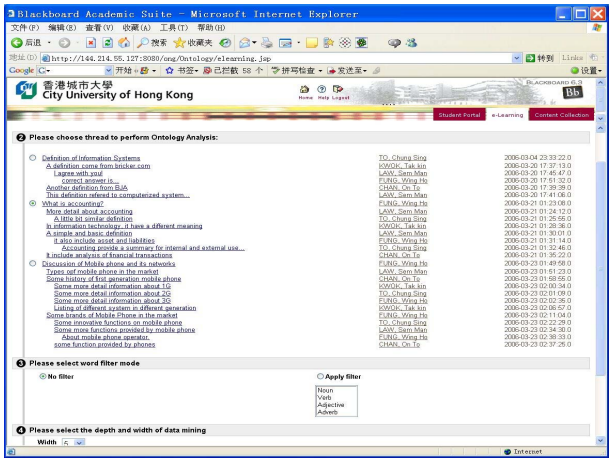


Fig. 7. Threads of Messages on Online Discussion Board

Figure 8 shows the enhanced Blackboard interface which provides access to the concept map tool installed on our development sever. The instructor can click the "Launch Concept Map Viewer" hyper-link to activate the "Ontology Saving and Concept Map Display" function as depicted in Figure 9. By clicking the hyper-link under the title Concept Map, a concept map will be displayed as shown in Figure 10. The corresponding owl statements generated by our system was shown in Figure 1. The user can also click the "Save Ontology File" hyper-link to save the owl file to a local disk. Figure 10 depicts the concept map about knowledge management, and the other concepts such as knowledge discovery, knowledge capture, intellectual capital, business management, etc. are the sub-concepts. For readability reason, stemming is not performed for our demonstration examples. As the size of each node on the concept map is fixed, some of the characters of the concept labels are truncated. When the user moves the mouse pointer over the node, all the words of the concept can be displayed. The number attached to a link connecting each pair of concepts shows the strength of the corresponding sub-concept/super-concept relationship. When a node at the second level is clicked, all the sub-concepts below the current node will be shown. For instance, when the instructor clicks the "knowledge capture" node (i.e., the node with the number "3" on the top right hand corner), the sub-concepts under this node

will be displayed as demonstrated in Figure 11. The number attached to the top right hand corner of a node indicates the number of levels below the current node.

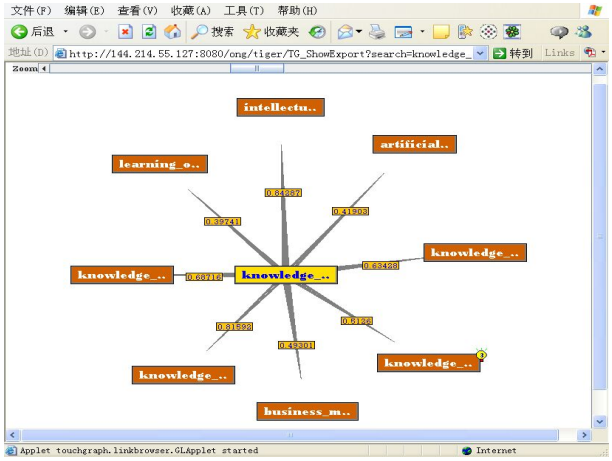


Fig. 10. An Automatically Generated Concept Map About "Knowledge Management"

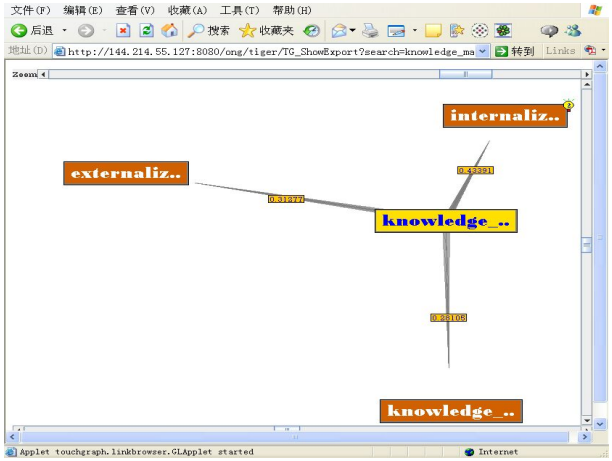


Fig. 11. The Display of Sub-concept "Knowledge Capture"

As a fuzzy domain ontology may contain hundreds of nodes, it may be difficult to display all the concepts in one single concept map. We adopt a linearization procedure [59] to generate the concepts map for each main concept when the number of nodes in an ontology exceeds 100. Figure 12 demonstrates the linear display of a fuzzy domain ontology which is constructed by scanning the full text of the first $1,000$ Web pages about "knowledge management" retrieved by using the Google Search API (`http://code.google.com/apis/soapsearch/`). After the user clicks the node "km_topology", the corresponding concept map will be displayed as in shown in Figure 13. It is obvious that the concept map as shown in Figure 13 is more noisy than the concept map generated from the more focused online class discussion shown in Figure 11. One of the main reasons is that there are many commercial spam embedded in the Web pages retrieved from Google.

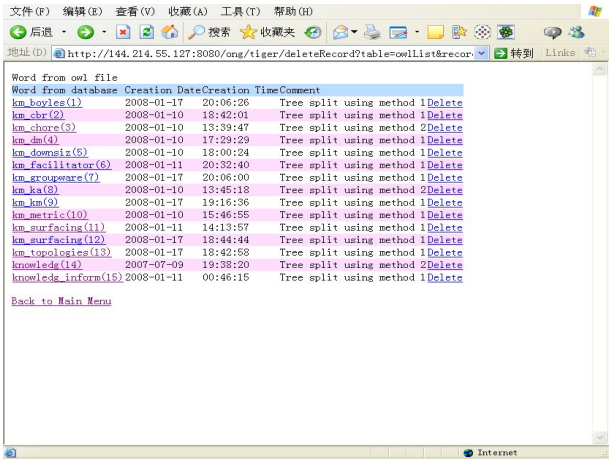


Fig. 12. The Linearized Concept Display

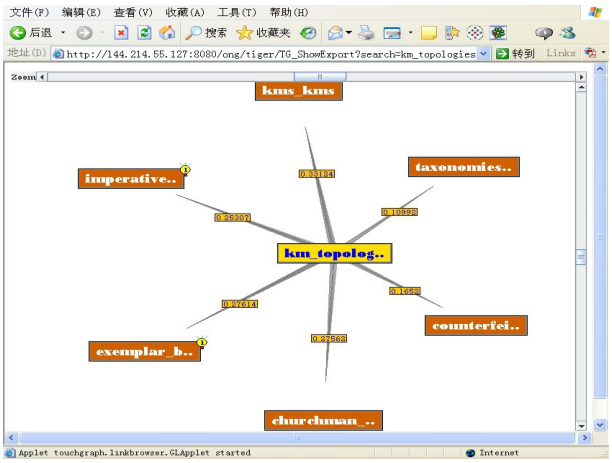


Fig. 13. A Concept Map Extracted From the Google Web Pages

Our prototype system is developed using Java (J2SE v 1.4.2), Java Server Pages (JSP) 2.1, and Servlet 2.5. For the implementation of singular value decomposition for our term space reduction, we employ the publicly available java toolkit called GAP (`http://gap.stat.sinica.edu.tw/Software/GAP/`). For the visualization of the concept maps, we develop our visualization module based on the java-based shareware TouchGraph (`http://sourceforge.net/projects/touchgraph`). Our prototype system is operated under Apache Tomcat 6.0.

## VII. System Evaluation

### A. Evaluation Metrics

We try to evaluate the automatically generated concept maps by comparing them with the maps developed by human experts. Our first evaluation metric is developed based on the Generalized Distance Ratio (GDR) method [32]. The GDR measure is the generalization of Langfield-Smith and Wirth's metric [22], and it has been widely used to quantitatively evaluate concept maps in the fields of education, operational

research and strategic management [32]. The GDR measure aims at comparing concept maps by using all the available information encoded in the maps. Specifically, the GDR measure considers three types of difference: (1) existence or non-existence of elements (nodes); (2) existence and non-existence of beliefs (arcs); (3) identical beliefs (arcs) held with differing strengths (i.e., the membership of our fuzzy relation $R_{CC}$).

Originally, the GDR measure has five parameters such as $\alpha, \beta, \gamma, \varepsilon,$ and $\delta$ to deal with different kinds of map comparisons [32]. To adopt the GDR to meet our specific map comparison requirements, we employ the following parameter values:

- $\alpha = 1$, represents no account for the values for nodes directly influencing themselves (self-influence);
- $\beta = \varepsilon = 1$ represent the weights of arcs in the unit interval $[0, 1]$;
- $\gamma = 1$, represents the normal way to interpret the matrix cells for which two maps cannot have an identical if the corresponding pair of nodes are not the same;
- $\delta = 0$, represents no special treatment to polarity change.

As a result, our adapted Generalized Distance Ratio measure (DR) becomes:

$$DR(A,B) = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} diff(i,j)}{p_c^2+2p_c(p_{u_A}+p_{u_B})+p_{u_A}^2+p_{u_B}^2-(p_c+p_{u_A}+p_{u_B})} \quad (23)$$

$$diff(i,j) = \begin{cases} 0 & \text{if i = j} \\ 1 & \text{if } (\mathrm{a_{ij}} \neq 0 \vee \mathrm{b_{ij}} \neq 0) \wedge \\ & (i \notin P_c \vee j \notin P_c) \\ |a_{ij} - b_{ij}| & \text{otherwise} \end{cases} \quad (24)$$

where:

- $A$ and $B$ are two extended adjacency matrices of size $p$,
- $a_{ij}$ (or $b_{ij}$) is the value of the $i$th row and $j$th column of ($A$ or $B$),
- $P_c$ is the set of nodes common to both maps,
- $p_c = |P_c|$ is the cardinality of the set $P_c$,
- $p_{uA}$ is the number of nodes unique to map $A$,
- $p_{uB}$ is the number of nodes unique to map $B$,
- $N_A$ and $N_B$ are the sets of nodes in the maps $A$ and $B$ respectively.

It yields the distance ratio, $DR$, of any two maps in the scale of [0, 1]. With reference to the example depicted in Figure 14, Table I and Table II depict the corresponding adjacency matrices. It should be noted that the measure DR can be applied to evaluate any pair of maps even if the number of nodes of these maps are not the same. The procedure of calculating the DR score Eq.(23) between map A and map B is illustrated as follows:

Step 1: Compute the number of unique nodes $p_{uA}$ and $p_{uB}$ for Map $A$ and Map $B$ respectively;

Step 2: Identify the set of common nodes $P_c$ and count its cardinality $p_c$;

Step 3: Determine the size of the adjacency matrices $p_{uA} + p_{uB} + p_c$;

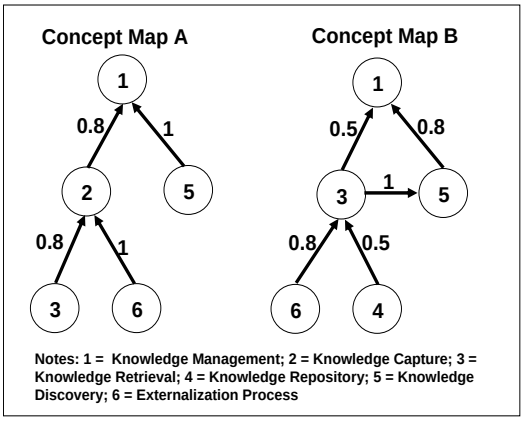


Fig. 14. An Example of Concept Map Comparison

| Nodes | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 0.8 | 0 | 0 | 1 | 0 |
| 2 | 0 | 0 | 0.8 | 0 | 0 | 1 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |

TABLE I
THE ADJACENCY MATRIX OF MAP A

| Nodes | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0.5 | 0 | 0.8 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0.5 | 1 | 0.8 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |

TABLE II
THE ADJACENCY MATRIX OF MAP B

Step 4: Represent Map $A$ and Map $B$ using the extended adjacency matrices. The weight of an arc is between 0 and 1.

Step 5: Calculate the $DR$ score using Eq.(23). With reference to our example, the following values are instantiated:

1) $p = 6$;
2) $P_c = \{1, 3, 5, 6\}, p_c = |P_c| = 4$;
3) $P_{uA} = \{2\}, p_{uA} = |P_{uA}| = 1$;
4) $P_{uB} = 4, p_{uB} = |P_{uB}| = 1$;
5) $N_A = \{1, 2, 3, 5, 6\}$;
6) $N_B = \{1, 3, 4, 5, 6\}$;
7) $p_c^2 + 2p_c(p_{uA} + p_{uB}) + p_{uB}^2 + p_{uC}^2 - (p_c + p_{uA}^2 + p_{uB}^2)$ = 16 + 16 + 1 + 1 - 6 = 28;
8) $\sum_{i=1}^{6}\sum_{j=1}^{6} diff(i,j)$ = (0 + 1 + 0.5 + 0 + 0.2 + 0) + (0 + 0 + 1 + 0 + 0 + 1) + (0 + 0 + 0 + 1 + 1 + 0.8) + (0 + 0 + 0 + 0 + 0 + 0) + (0 + 0 + 0 + 0 + 0 + 0) + (0 + 0 + 0 + 0 + 0 + 0 + 0) = 6.5
9) $DR = 6.5/28 = 0.232$

Notes:

$i$=1, $j$=5, $diff(i,j)$ = $|a_{15} - b_{15}| = |1 - 0.8| = 0.2$;

$i$=1, $j$=4, $diff(i,j)$ = $|a_{14} - b_{14}|$ = 0, because $a_{14}$ = 0 and $b_{14}$ = 0 even node 4 is absent in Map $A$;

$i$=1, $j$=2, $diff(i,j)$ =1, because $a_{12} \neq 0$ and node 2 is absent in Map $B$;

$i$=1, $j$=1, $diff(i,j)$ =0.

Step 6: The similarity between the Map $A$ and the Map $B$ is the dural of the $DR$ score; therefore $Sim(A,B) = 1 - DR = 0.768$.

We also employ standard measures such as recall, precision, and the F-measure developed in the field of IR [44] to evaluate the concept maps. In particular, we develop ontology recall $Ont\_Recall$, ontology precision $Ont\_precision$, and ontology F-measure $Ont\_F$ as follows:

$$Node\_Recall = \frac{|N_{M_E} \cap N_{M_S}|}{|N_{M_E}|} \tag{25}$$

$$Node\_Precision = \frac{|N_{M_E} \cap N_{M_S}|}{|N_{M_S}|} \tag{26}$$

$$Link\_Recall = \frac{|L_{M_E} \cap L_{M_S}|}{|L_{M_E}|} \tag{27}$$

$$Link\_Precision = \frac{|L_{M_E} \cap L_{M_S}|}{|L_{M_S}|} \tag{28}$$

$$\begin{aligned} Ont\_Recall = \quad & \omega_R \times Node\_Recall + \\ & (1 - \omega_R) \times Link\_Recall \end{aligned} \tag{29}$$

$$\begin{aligned} Ont\_Precision = \quad & \omega_P \times Node\_Precision + \\ & (1 - \omega_P) \times Link\_Precision \end{aligned} \tag{30}$$

$$F_\eta = \frac{(1+\eta^2) Precision \times Recall}{\eta^2 Precision + Recall} \tag{31}$$

where $N_{M_E}$ and $N_{M_S}$ represent the set of nodes found from the concept map created by human experts and that generated by our system respectively. Similarly, $L_{M_E}$ and $L_{M_S}$ are the set of links encoded on the concept map drawn by human experts and the concept map generated by our system respectively. In fact, the set of links can easily be extracted from the adjacency matrices like those depicted in Table I and Table II. In particular, only the upper half or the lower half of each matrix needs to be traversed to construct a link set. For instance, $L_A = \{l_{ij} \in L : a_{ij} > 0\}$, whereas $L$ is the set of all possible links encoded on the maps. The parameter $\omega_R$ is used to compute the ontology recall based on a weighted sum of the node recall and link recall respectively. Similarly, $\omega_P$ is used to tune the ontology precision measure. For the experiments presented in this paper, we adopt $\omega_R = \omega_P = 0.5$. The standard F-measure is shown in Eq.(31) [54]. If we assume that precision is as important as recall (i.e. $\eta = 1$), the ontology F-measure $Ont\_F$ is reduced to:

$$Ont\_F = \frac{2 \times Ont\_Precision \times Ont\_Recall}{Ont\_Precision + Ont\_Recall} \tag{32}$$

### B. Benchmark Tests

For the initial experiments, we used a benchmark corpus developed in the Text REtrieval Conference (TREC) [42] to evaluate our system. In TREC, many TREC Topics were developed to represent distinct information topics (domains). Corresponding to the TREC topics are several collections of documents used to test the effectiveness of IR systems. The TREC-AP, which comprises the Associated Press (AP) newswires covering the period from 1988 to 1990, is one of the benchmark corpora used. In our experiments, we used a TREC topic description and at most five relevant TREC-AP documents associated with the particular topic to simulate the online messages generated from an e-Learning platform. A human expert (a post-doctoral researcher in the field of banking and finance) was recruited to read the topic description as well as the associated documents so that she could draw a concept map illustrating the main concepts and propositions for each of the TREC topic. These concept maps become the benchmark for comparison with the system generated concept maps based on the metrics developed in Section VII-A. For each experimental run, we manipulated different system parameters to test different aspects of our system. We selected TREC topics 1 to 10 and topics 41 to 50 for our experiments since each topic has at least 5 relevant documents. The sample of TREC topic 49 (Who's working with Supercomputers) is shown in Appendix A and the corresponding first level concept map generated by our system is depicted in Figure 15.

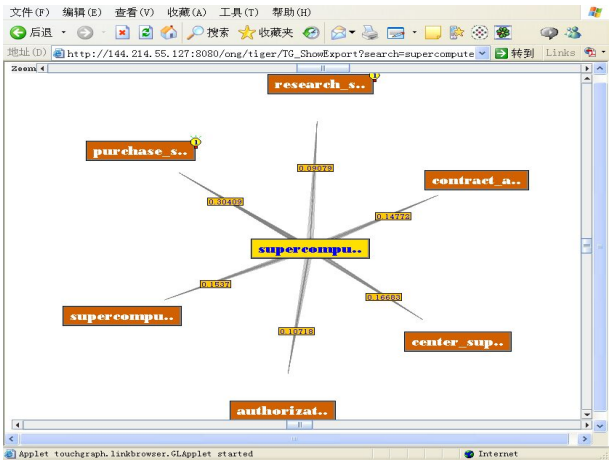


Fig. 15. First Level Concept Map Generated from TREC Topic 49

Experiment 1

The purpose of the first experiment is to test the effectiveness of the concept extraction/filtering thresholds. We used the BMI method Eq.(2) for concept extraction and the standard fuzzy conjunction operation Eg.(14) for fuzzy relation extraction. Other parameters included $\beta = 0.672$ [25] and $\lambda = 0.093$. The noun phrase patterns "Noun Noun" were used for all the experiments discussed in this paper. We used other five domains (e.g., entertainment, education, humanity,

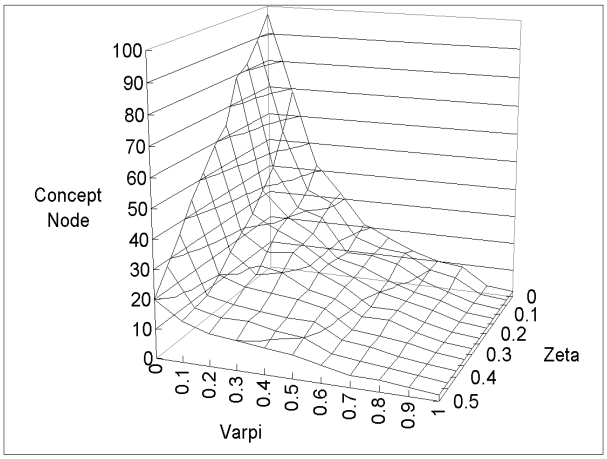


Fig. 16. Average Number of Concepts Generated by Controlling $\zeta$ and $\varpi$

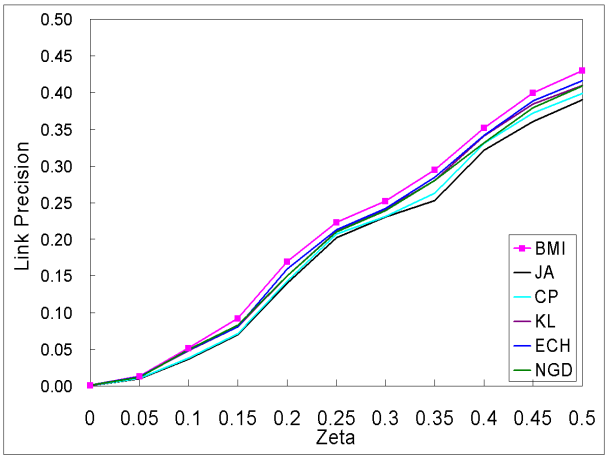


Fig. 18. The Relative Link Precision of Various Concept Extraction Methods

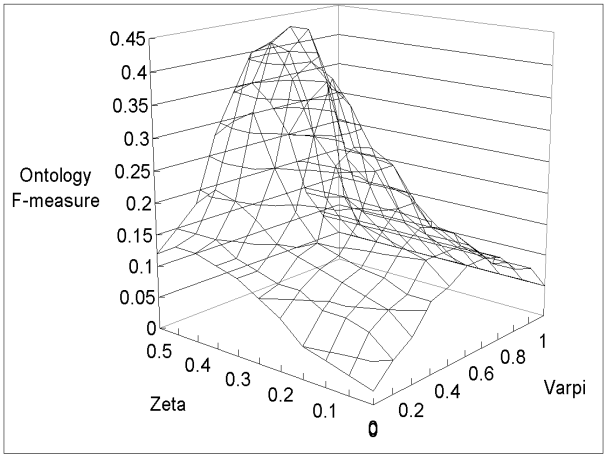


Fig. 17. The Average F-measure by Tuning $zeta$ and $varpi$

sport, and arts) as the basis to compute the concept relevance scores during concept filtering. Each domain consists of $1,000$ Web pages retrieved from Google via the Google Search API. When the domain frequency $Dom(c_i, D_k)$ was calculated for the TREC-AP domain, we converted the document basis to 1,000 as well. The average number of concept nodes generated and the average ontology F-measure achieved over the twenty TREC domains under various combinations of $\zeta$ and $\varpi$ are plotted in Figure 16 and Figure 17 respectively. As shown in Figure 16, when both $\zeta \geq 0.2$ and $\varpi \geq 0.4$ were applied, the number of concept nodes generated by our system would be reduced dramatically. It indicates that our concept filtering mechanism can work effectively. From Figure 17, it is shown that the best ontology F-measure could be achieved if $\zeta$ and $\varpi$ are set to the ranges $[0.35, 0.45]$ and $[0.4, 0.6]$ respectively. The reason is that most noisy concepts will be filtered out under such a combination.

Experiment 2

The objective of this experiment is to evaluate the effectiveness of different concept extraction methods such as BMI, JA, CP, KL, ECH, and NGD illustrated in Section V-A. When different methods are applied, the underlying terms and the term weights $\mu_{c_i}(t)$ associated with a concept may be different. Such a difference can be realized when we apply Eq.(22) or Eq.(14) to compute the fuzzy relations between concepts because both of the metrics will compare the concepts based on their underlying semantics (e.g., the composing terms and their weights). We adopted the same system parameters as in experiment one. The average link precisions achieved by various concept extraction methods under different extraction threshold values $\zeta = [0, 0.5]$ are plotted in Figure 18. In general, the link precision is improved when higher concept extraction threshold is used because less noisy terms will be used to construct the corresponding concept vectors. As shown in Figure 18, the BMI method outperforms the other methods in terms of average link precision at all $\zeta$ levels.

Experiment 3

We also examined the effectiveness of Eq.(22) and Eq.(14) which were used to estimate the strength of a concept specialization relation $\mu_{R_{CC}}(c_x, c_y)$ given any two concepts $c_x, c_y$. In this experiment, we used the BMI concept extraction method and we set the parameters $\zeta = 0.431$ and $\varpi = 0.512$; other parameters were not changed except the two relation extraction methods. In the first run, we used the standard fuzzy conjunction operator Eq.(14) for concept map generation, and then we employed the same set of parameters to invoke the structural similarity SSIM method Eq.(22). The parameters used for SSIM were: $Q_1 = 0.026, Q_2 = 0.459, Q_3 = 0.344$. A topic-by-topic comparison in terms of ontology precision, ontology recall, ontology F-measure, and DR are tabulated in Table III and Table IV respectively. The second and the third columns refer to the number of concept nodes and concept relations generated by the system. By testing the hypotheses: $H_{Null} : \mu_{SSIM} - \mu_{Fuzzy} = 0$ and $H_{Alternative} : \mu_{SSIM} - \mu_{Fuzzy} > 0$ with paired one-tail $t$-test on the F-measure scores obtained from the 20 TREC-AP topics, the null hypothesis is rejected ($t(19) = 3.067, p < 0.01$). Therefore, it is confirmed that the SSIM method Eq.(22) for concept relation extraction is more effective than the method using

standard fuzzy conjunction operator Eq.(14). As can be seen, the average distance between the maps generated by our system and the maps drawn by the domain expert is 0.285 only. It means that the concept maps produced by our system can really represent the domain knowledge as perceived by the domain expert.

| Topic | Node | Link | Ontology Recall | Ontology Precision | Ontology F-measure | DR |
|---|---|---|---|---|---|---|
| 1 | 25 | 116 | 0.763 | 0.474 | 0.585 | 0.267 |
| 2 | 27 | 95 | 0.733 | 0.568 | 0.640 | 0.215 |
| 3 | 22 | 91 | 0.691 | 0.615 | 0.651 | 0.199 |
| 4 | 18 | 90 | 0.691 | 0.522 | 0.595 | 0.244 |
| 5 | 16 | 88 | 0.765 | 0.409 | 0.533 | 0.324 |
| 6 | 24 | 115 | 0.673 | 0.287 | 0.402 | 0.418 |
| 7 | 27 | 131 | 0.688 | 0.419 | 0.521 | 0.297 |
| 8 | 18 | 93 | 0.754 | 0.527 | 0.620 | 0.213 |
| 9 | 19 | 87 | 0.712 | 0.540 | 0.614 | 0.254 |
| 10 | 22 | 98 | 0.731 | 0.388 | 0.507 | 0.387 |
| 41 | 16 | 89 | 0.634 | 0.292 | 0.400 | 0.318 |
| 42 | 24 | 81 | 0.625 | 0.370 | 0.465 | 0.329 |
| 43 | 23 | 85 | 0.711 | 0.694 | 0.702 | 0.157 |
| 44 | 19 | 96 | 0.600 | 0.375 | 0.462 | 0.330 |
| 45 | 20 | 97 | 0.727 | 0.412 | 0.526 | 0.279 |
| 46 | 22 | 109 | 0.754 | 0.395 | 0.518 | 0.303 |
| 47 | 18 | 89 | 0.736 | 0.315 | 0.441 | 0.321 |
| 48 | 19 | 85 | 0.667 | 0.353 | 0.462 | 0.362 |
| 49 | 16 | 101 | 0.683 | 0.277 | 0.394 | 0.346 |
| 50 | 15 | 87 | 0.806 | 0.333 | 0.472 | 0.356 |
| Avg. | 20.5 | 96.15 | 0.707 | 0.428 | 0.526 | 0.296 |

TABLE III
CONCEPT MAP GENERATION USING STANDARD FUZZY CONJUNCTION OPERATOR

| Topic | Node | Link | Ontology Recall | Ontology Precision | Ontology F-measure | DR |
|---|---|---|---|---|---|---|
| 1 | 25 | 117 | 0.748 | 0.464 | 0.572 | 0.272 |
| 2 | 27 | 97 | 0.721 | 0.536 | 0.615 | 0.231 |
| 3 | 22 | 90 | 0.704 | 0.633 | 0.667 | 0.190 |
| 4 | 18 | 93 | 0.676 | 0.495 | 0.571 | 0.258 |
| 5 | 16 | 85 | 0.766 | 0.424 | 0.545 | 0.316 |
| 6 | 24 | 116 | 0.633 | 0.267 | 0.376 | 0.429 |
| 7 | 27 | 125 | 0.738 | 0.472 | 0.576 | 0.270 |
| 8 | 18 | 90 | 0.769 | 0.556 | 0.645 | 0.199 |
| 9 | 19 | 88 | 0.727 | 0.545 | 0.623 | 0.251 |
| 10 | 22 | 93 | 0.750 | 0.419 | 0.538 | 0.367 |
| 41 | 16 | 86 | 0.683 | 0.329 | 0.444 | 0.302 |
| 42 | 24 | 74 | 0.687 | 0.446 | 0.541 | 0.289 |
| 43 | 23 | 82 | 0.735 | 0.744 | 0.739 | 0.131 |
| 44 | 19 | 99 | 0.583 | 0.354 | 0.440 | 0.342 |
| 45 | 20 | 86 | 0.764 | 0.488 | 0.596 | 0.243 |
| 46 | 22 | 101 | 0.772 | 0.436 | 0.557 | 0.283 |
| 47 | 18 | 86 | 0.763 | 0.337 | 0.468 | 0.310 |
| 48 | 19 | 83 | 0.689 | 0.373 | 0.484 | 0.351 |
| 49 | 16 | 93 | 0.732 | 0.323 | 0.448 | 0.324 |
| 50 | 15 | 82 | 0.806 | 0.354 | 0.492 | 0.345 |
| Avg. | 20.5 | 93.25 | 0.722 | 0.449 | 0.547 | 0.285 |

TABLE IV
CONCEPT MAP GENERATION USING SSIM

### C. Field Tests

Field tests were conducted to verify the quality of the concept maps generated by our system. The subjects were the postgraduate students attending a Knowledge Management course. These subjects learnt about concept mapping in their classes. At the end of a lecture, subjects were told to reflect the main concepts they learnt from the class by writing short messages on an online discussion forum. The time given to them to write the messages was limited to ten minutes for each class. After the subjects had finished their reflection, the concept map generation tool was invoked to automatically construct the concept maps representing the group's perception about the concepts covered in the lecture. We employed the BMI method for concept extraction and the SSIM method for relation extraction. Other system parameters were the same as those used in experiment three. Each subject was given another 10 minutes to browse through the concept maps generated by the system, and then they would answer a questionnaire. Our questionnaire was developed based on the instrument employed by [6]. It included the assessment of the following factors:

- Accuracy - Whether the concepts and relationships shown at the taxonomy are correct;
- Cohesiveness - Whether each concept at the taxonomy is unique and not overlapped with one another;
- Isolation - Whether the concepts at the same level are distinguishable and not subsume one another;
- Hierarchy - Whether the taxonomy is traversed from broader concepts at the higher levels to narrow concepts at the lower level;
- Readability - Whether the concepts at all levels are easy to be comprehended by human;

A five point semantic differential scale from very good (5), good (4), average (3), bad (2), to very poor (1) is used to measure the dependent variables. In general, a score close to 5 indicates that the automatically generated concept map is with good quality and it can reflect what the group perceived about the subject topic. The results of the field tests are shown in Table V. The second column indicates the number of subjects involved in a field test, the third and the fourth columns show the number of concepts nodes and links automatically generated by the system, and the fifth column shows the time spent on generating the concept maps. The overall mean scores for accuracy, cohesiveness, isolation, hierarchy, and readability are $4.23$, $4.22$, $4.15$, $4.31$, and $3.95$ respectively. For most of the dependent variables, the overall mean score is above $4$ except the readability issue. The reason for a bit lower score in readability may be that our programmer used a fixed size rectangle to represent concept node. As a noun phrase (two words) often cannot fit into such a rectangle, subjects might not know what the concept is about from a glance. As the Tough Graph shareware supports the display of variable sized nodes, it is easy for us to improve the readability of the concept maps in the future. The time taken to generate the concept map (including the underlying OWL statements) on our Web server varied from $1.3$ to $1.8$ minutes. This result indicates that it is feasible for instructors to invoke such a tool to analyze students' progress on the fly.

## VIII. CONCLUSIONS AND FUTURE WORK

With the increasing number of online messages generated from interactive e-Learning environments, instructors are often

| Lecture | Subject | Node | Link | Time | Accuracy | | Cohesiveness | | Isolation | | Hierarchy | | Readability | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Mean | STD | Mean | STD | Mean | STD | Mean | STD | Mean | STD |
| Knowledge Management | 20 | 16 | 61 | 1.5 | 4.20 | 0.871 | 4.25 | 0.698 | 4.05 | 0.865 | 4.41 | 0.663 | 3.93 | 0.831 |
| Knowledge Discovery | 19 | 17 | 63 | 1.6 | 4.21 | 0.885 | 4.22 | 0.673 | 4.15 | 0.763 | 4.32 | 0.822 | 4.01 | 0.642 |
| Knowledge Sharing | 22 | 21 | 79 | 1.8 | 4.25 | 0.626 | 4.17 | 0.759 | 4.16 | 0.833 | 4.28 | 0.682 | 3.87 | 0.693 |
| Knowledge Capture | 20 | 18 | 55 | 1.5 | 4.33 | 0.678 | 4.35 | 0.622 | 4.29 | 0.654 | 4.29 | 0.731 | 4.13 | 0.834 |
| Knowledge Application | 18 | 16 | 59 | 1.3 | 4.17 | 0.654 | 4.09 | 0.827 | 4.11 | 0.846 | 4.23 | 0.693 | 3.82 | 0.675 |
| Average | | 17.60 | 63.40 | 1.54 | 4.23 | 0.743 | 4.22 | 0.716 | 4.15 | 0.792 | 4.31 | 0.718 | 3.95 | 0.735 |

TABLE V
THE RESULTS OF FIELD TESTS

overwhelmed and hence adaptive teaching and learning is difficult. This paper illustrates a novel concept map generation technique which is underpinned by a context-sensitive text mining method and a fuzzy domain ontology extraction algorithm. The proposed mechanism can automatically construct concept maps based on the messages posted to an online discussion board. By providing such a tool to an e-Learning environment, instructors can quickly identify the learning status of their students, and hence more suitable pedagogy can be developed for the subsequent lessons. Our initial experimental results show that the accuracy and the quality of the automatically generated concept maps is promising. Future work involves a larger scale of field test against our automated concept map generation mechanism. Other text mining methods will also be explored to improve our fuzzy domain ontology extraction method.

## IX. ACKNOWLEDGMENTS

The work reported in this paper has been funded in part by the Australian Research Council (ARC) Discovery grant (DP0556455). This research work is also supported in part by the e-Learning research grant (Project No.: 6980081-680) of the City University of Hong Kong, and the UK's Engineering and Physical Sciences Research Council (EPSRC) grant (EP/E002145/1).

## REFERENCES

[1] Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel. An algorithm that learns what's in a name. *Machine Learning*, 34:211–231, 1999.

[2] Johanna Bucur. HyWrite: writing in hypermedia eLearning environments. In *Proceedings of the seventeenth conference on Hypertext and hypermedia (HYPERTEXT'06)*, pages 45–48. ACM Press, 2006.

[3] C. Calvi. Navigation and disorientation: A case study. *Journal of Educational Multimedia and Hypermedia*, 6(3–4):305–320, 1997.

[4] Shan Chen, Damminda Alahakoon, and Maria Indrawan. Background knowledge driven ontology discovery. In *Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service*, pages 202–207, 2005.

[5] Yu-Liang Chi. Elicitation synergy of extracting conceptual tags and hierarchies in textual document. *Expert Systems with Applications*, 32(2):349–357, 2007.

[6] S. Chuang and L. Chien. Taxonomy generation for text segments: A practical web-based approach. *ACM Transactions on Information Systems*, 23(4):363–396, 2005.

[7] R.L. Cilibrasi and P.M.B. Vitanyi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–383, 2007.

[8] P. Cimiano, A. Hotho, and S. Staab. Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 24:305–339, 2005.

[9] The World Wide Web Consortium. Web Ontology Language, 2004. Available from `http://www.w3.org/2004/OWL/`.

[10] Michael Dittenbach, Helmut Berger, and Dieter Merkl. Improving domain ontologies by mining semantics from text. In *Proceedings of the First Asia-Pacific Conference on Conceptual Modelling (APCCM2004)*, pages 91–100, 2004.

[11] G. Forsythe, M. Malcolm, and C. Moler. *Computer methods for mathematical computations*. Prentice Hall, Englewood Cliffs, 1977.

[12] George Furnas, Scott Deerwester, Susan T. Dumais, Thomss K. Landauer, Richard Harshman, Lynn A. Streeter, and Karen E. Lochbaum. Information retrieval using a singular value decomposition model of latent semantic structure. In Yves Chiaramella, editor, *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 465–480, Grenoble, France, 1988. ACM Press.

[13] I. Ganchev, M. O'Droma, and R. Andreev. Functionality and SCORM-compliancy Evaluation of eLearning Tools. In *Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies*, pages 467–469. IEEE Computer Society, 2007.

[14] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[15] Z. Harris. *Mathematical Structures of Language*. Wiley, New York, 1968.

[16] Larry Howard, Zsolt Remenyi, and Gabor Pap. Adaptive blended learning environments. In *Proceedings of the 9th International Conference on Engineering Education*, pages 11–16, 2006.

[17] Hongyan Jing and Evelyne Tzoukermann. Information retrieval based on context distance and morphology. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Language Analysis, pages 90–96, 1999.

[18] I. T. Jolliffe. Principal component analysis. In *Principal Component Analysis*. Springer Verlag, New York, 1986.

[19] D.H. Jonassen, K. Beissner, and M. Yacci. *Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge*. Lawrence Erlbaum Associates Publishers, Hillsdale: NJ, 1993.

[20] J. Kay and S. Holden. Automatic extraction of ontologies from teaching document metadata. In *Proceedings of the 2002 International Conference on Computers in Education*, pages 1555–1556. IEEE Computer Society, 2002.

[21] D. Koller and M. Sahami. Hierarchically classifying documents using very few words. In Douglas H. Fisher, editor, *Proceedings of ICML-97, 14th International Conference on Machine Learning*, pages 170–178, Nashville, Tennessee, 1997. Morgan Kaufmann Publishers, San Francisco, California.

[22] K. Langfield-Smith and a. Wirth. Measuring differences between cognitive maps. *Journal of the Operational Research Society*, 43(12):1135–1150, 1992.

[23] Raymond Y.K. Lau, Albert Y.K. Chung, Dawei Song, and Qiang Huang. Towards Fuzzy Domain Ontology Based Concept Map Generation for e-Learning. In H. Leung et. al., editor, *Proceedings of the Sixth International Conference on Web-based Learning (ICWL'07)*, volume 4823 of *Lecture Notes in Computer Science*, pages 90–101, Edinburgh, UK, 2008. Springer.

[24] R.Y.K. Lau. Context-Sensitive Text Mining and Belief Revision for Intelligent Information Retrieval on the Web. *Web Intelligence and Agent Systems An International Journal*, 1(3-4):1–22, 2003.

[25] R.Y.K. Lau. Fuzzy Domain Ontology Discovery for Business Knowledge Management. *IEEE Intelligent Informatics Bulletin*, 8(1):29–41, 2007.

[26] R.Y.K. Lau, Yuefeng Li, and Yue Xu. Mining Fuzzy Domain Ontology from Textual Databases. In *Proceedings of the 2007 IEEE/WIC/ACM*

*International Conference on Web Intelligence*, Silicon Valley, CA, November 2–5 2007. IEEE Press. On CR-ROM.

[27] Chang-Shing Lee, Zhi-Wei Jian, and Lin-Kai Huang. A fuzzy ontology and its application to news summarization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 35(5):859–880, 2005.

[28] D. Levy. Users and interaction track: memex and hypertext: To grow in wisdom: vannevar bush, information overload, and the life of leisure. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pages 281–286, 2005.

[29] Yuefeng Li and Ning Zhong. Mining ontology for automatically acquiring web user information needs. *IEEE Transactions on Knowledge and Data Engineering*, 18(4):554–568, 2006.

[30] A. Maedche, V. Pekar, and S. Staab. Ontology learning part onełon discovering taxonomic relations from the web. In N. Zhong, J. Liu, and Y. Yao, editors, *Web Intelligence*, pages 3–24. Springer, 2003.

[31] Alexander Maedche and Steffen Staab. Ontology learning. In *Handbook on Ontologies*, pages 173–190. 2004.

[32] L. Markoczy and J. Goldberg. A method for eliciting and comparing causal maps. *Journal of Management*, 21(2):305–309, 1995.

[33] G. A. Miller, Beckwith R., C. Fellbaum, D. Gross, and K. J. Miller. Introduction to wordnet: An on-line lexical database. *Journal of Lexicography*, 3(4):234–244, 1990.

[34] Michele Missikoff and Francesco Taglino. An ontology-based platform for semantic interoperability. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, pages 617–634. Springer, 2004.

[35] Christine A. Montgomery. Concept extraction. *American Journal of Computational Linguistics*, 8(2):70–73, 1982.

[36] Roberto Navigli, Paola Velardi, and Aldo Gangemi. Ontology learning and its application to automated terminology translation. *IEEE Intelligent Systems*, 18(1):22–31, 2003.

[37] J.D. Novak and D.N. Gowin. *Learning How to Learn*. Cambridge Univ. Press, Cambridge, U.K., 1984.

[38] C. Papatheodorou, A. Vassiliou, and B. Simon. Discovery of ontologies for learning resources using word-based clustering. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA'02)*, pages 324–326, 2002.

[39] Patrick Perrin and Frederick Petry. Extraction and representation of contextual information for knowledge discovery in texts. *Information Sciences*, 151:125–152, 2003.

[40] M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[41] P.F.W. Preece. Mapping cognitive structure: A comparison of methods. *Journal of Educational Psychology*, 68:1–8, 1976.

[42] S. Robertson and I. Soboroff. The TREC 2001 Filtering Track Report. In E.M. Voorhees and D.K. Harman, editors, *Proceedings of the tenth Text REtrieval Conference (TREC-10)*, pages 26–37, Gaithersburg, Maryland, November 13–16 2001. Department of Commerce, NIST. Available from `http://trec.nist.gov/pubs/trec10/t10_proceedings.html`.

[43] G. Salton. Full text information processing using the smart system. *Database Engineering Bulletin*, 13(1):2–9, March 1990.

[44] G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, New York, 1983.

[45] M. Sanderson and B. Croft. Deriving concept hierarchies from text. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 206–213. ACM, 1999.

[46] Hinrich Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124, 1998.

[47] Satoshi Sekine, Jeremy J. Carroll, Sofia Ananiadou, and Jun'ichi Tsujii. Automatic learning for semantic collocation. In *Proceedings of the third Conference on Applied Natural Language Processing*, pages 104–110, Trento, Italy, March 31–April 3 1992. Association for Computational Linguistics.

[48] C. Shannon. A mathematical theory of communication. *Bell System Technology Journal*, 27:379–423, 1948.

[49] Martyn Sloman. *The e-learning revolution : how technology is driving a new training paradigm*. AMACOM, 2002.

[50] G. Smith. *Computers and Human Language*. Oxford University Press, New York, New York, 1991.

[51] M. H. Soong, H. C. Chan, B. C. Chua, and K. F. Loh. Critical succes factors for on-line course resources. *Computers & Education*, 36(2):101–120, 2001.

[52] Mark A. Stairmand. Textual context analysis for information retrieval. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 140–147, 1997.

[53] Quan Thanh Tho, Siu Cheung Hui, Alvis Cheuk M. Fong, and Tru Hoang Cao. Automatic fuzzy ontology generation for semantic web. *IEEE Transactions on Knowledge and Data Engineering*, 18(6):842–856, 2006.

[54] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, London, United Kingdom, 1979.

[55] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[56] Christopher A. Welty. Ontology research. *AI Magazine*, 24(3):11–12, 2003.

[57] J. H. Wilkinson and C. Reinsch. *Handbook for Automatic Computation. Vol. II Linear Algebra*. Springer-Verlag, New York, 1971.

[58] Y. Wu and X. Chen. elearning assessment through textual analysis of class discussions. In *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies*, pages 388–390. IEEE Computer Society, 2005.

[59] J. Yeh, C. Wu, M. Chen, and L. Yu. Automated alighment and extraction of a bilingual ontology for cross-language domain-specific applications. *Computational Linguistics and Chinese Language Processing*, 10(1):35–52, 2005.

[60] L. A. Zadeh. Fuzzy sets. *Journal of Information and Control*, 8:338–353, 1965.

## Appendix A - TREC Topic 49 Description

```
<top>
<head> Tipster Topic Description
<num> Number: 049
<dom> Domain: Science and Technology
<title> Topic: Who’s working with Supercomputers
<desc> Description:
Document must identify an organization involved in the operation,
programming or purchase of a supercomputer.
<narr> Narrative:
To be relevant, a document must identify one of the following: a
supercomputing center, a supercomputer purchase, a supercomputer export
authorization, or the granting of a contract to a company known to perform
supercomputer support services.
<con> Concept(s):
1. Supercomputer, Cray, IBM 3090
2. Contract, authorization, purchase, sale, establish
3. Research
<fac> Factor(s):
<def> Definition(s):
Supercomputer- the most powerful computers available, typically
consisting of multiple processors optimized to execute in the most
efficient manner possible.
</top>
```