A Search Mechanism Based on Ontology Technology for Students in Elementary School

DOWMING YEH, SU-LING HSIANG
Department of Software Engineering
National Kaohsiung Normal University
116 Hou-Ping First Road, Kaohsiung
TAIWAN, REPUBLIC OF CHINA
dmyeh@nknu.edu.tw, hsl1008@gmail.com

Abstract: - Searching information on the Web has become part of everyone's daily activity in the modern society. The Web brings much convenience into our lives; however, with the amount of information growing in a staggering speed, how to sieve out useful information and knowledge effectively has become an important skill. In order to obtain a better search result, one needs to know how to form appropriate key words and phrases for search engines. However, elementary school students don't have sufficient abilities of sub-cognition, and they often form a search string in a natural language which contains few keywords in a long sentence; therefore, the search results do not often satisfy their need. This study addresses this problem by introducing an agent to parse the search string, align synonyms, sift out keywords, and include hyponyms in the intended field. Knowledge in a specific learning unit is represented with Ontology technology and transformed into keywords and associated structure. This study also integrates Google API and establishes specific function characteristics to eliminate unrelated files to conduct the search. The result shows that the prototype system can provide pupils with more accurate results.

Key-Words: search engine, empirical study, natural language, agent, ontology

1 Introduction

Since the era of Internet dawns, the Ministry of Education in Taiwan emphasizes that students should enhance their ability in information technology. With more and more information and knowledge published on the World Wide Web (WWW), the ability to search for information on the Web is certainly an important skill for students. Teachers in the elementary schools, therefore, often encourage their students to fulfill assignments by searching for related information with course contents. It is challenging to students since the results of searching the WWW usually contain many Web pages from different sources with various levels of quality, and students must learn how to sieve out the desired information. It is considered a good exercise for students to test their mental maturity and knowledge ability as well as foster their characters such as patience.

To search for the desired information, students have to resort to search engines. Although there are many powerful search engines available today, appropriate search strategies are often necessary to accomplish a successful or an efficient search. Even a simple search would need to form a keyword or a combination of keywords [9]. Researches show that many students in elementary schools fail to accomplish their searching processes because of choosing the inefficient browsing strategies or techniques [6]. These failures are often attributed to factors such as not knowing the right steps of using search engines and not being able to choose the correct keywords or use the suitable Boolean operations to form a search string. While the learning of search engines may be achieved with sufficient training, most students, due to their semantic deficiency, face difficulties in forming proper keywords or keywords with more specific meaning. As a result, many outcomes of their searches contain too many pages covering too broad a range for them to survey. In fact, many students form their search strings in their natural language. That is, they submit the search as though they are asking questions of their teachers. The same behavior is reported by Schacter et al [8]. Their research focused on how five graders and six graders solve their mission through Internet. In the process, 32 students did not adopt the Boolean expressions or specific keywords. Only a few students used the synonyms to search. Meanwhile, 30 students used the whole question sentence to search for their data. For example, they key in "What are the three common crimes in California?" as their keywords.

Previous search engines match keywords by character comparison and ignore the meaning and synonym of the keywords. Consequently, the search results fail to meet the requirement of users even though users provide keywords with similar meanings. Although nowadays search engines such as Google and Yahoo Website provide similar terms as well as search results to alleviate such problems, these offerings are still quite limited in their semantic range.

This paper describes a prototype of a search mechanism base on ontology which tries to solve the problems of forming proper search strings for students in elementary school. First of all, content in a specific learning unit is represented with ontology technology, transformed into keywords and associated structure, and stored in a repository. Then an agent is called to parse the search string which could be in a form of natural language and separate verbs and nouns from the string. Keywords are identified by aligning synonyms in the ontology structure. Moreover, hyponyms are included to facilitate more specific search. This prototype integrates with Google API and establishes specific function characteristics to eliminate unrelated files in specific formats. Finally, we compare the prototype with the traditional search engine and the result shows that the prototype system can provide students with more accurate results.

2 Ontology

Ontology comes from philosophy. The original meaning is the existing essences. In the case of computer science, ontology is a formal, explicit specification of a shared conceptualization [2]. It mainly designates the common cognition in some application domains and formal and specific definitions that clearly and precisely express conceptualized matters and represented clearly between computer and human [1]. Ontology can lower the barriers made by different terms and viewpoints [10].

Ontology has been applied to many research fields. The semantics knowledge from different fields can be gathered from the knowledge of ontology. Ontology can be divided into four kinds to meet different purposes [3]:

1. Top-level ontology mainly describes some general concepts, for example space, time, object, event and action. These general concepts do not relate to specific problem domains.
2. Domain ontology and task ontology describe some vocabularies related to specific domains such as medicine and automobile or specific task and activity such as diagnosis and sale. Domain knowledge corresponds to knowledge of an expert, recording concepts in the domain [4]. In this study, the domain ontology describes the content of a learning unit in the social study course of the elementary curriculum.
3. Application ontology describes concepts that are related to roles played by domain entities when performing certain activities.
4. Fig. 1 shows the relationships among the four kinds of ontology.

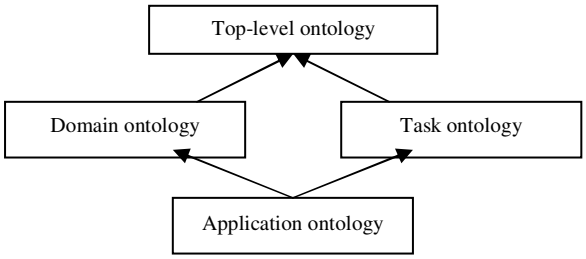

Fig. 1 Kinds of ontology

3 Ontology-based Search Mechanism

Our prototype is developed in Windows XP Professional with Microsoft SQL 2005 and ASP.NET 2.0 for VB. Fig. 2 illustrates the major components of the prototype system: Ontology Base, keyword repository, Chinese Knowledge and Information Processing unit, the Google search engine and an intelligent agent that coordinates these components and interacts with users. The operation of the whole system starts with users inputting a search string in natural language. Then CKIP API is called to analyze the search string and break it into verbs and nouns. Nouns are matched with words in the keyword repository to produce proper keywords and hyponyms. In the final stage, the API of Google is called to search for the desired information and the searching results are displayed to users.

The Chinese Knowledge and Information Processing (CKIP) is developed by the Academia Sinica (http://ckipsvr.iis.sinica.edu.tw/). The system can be accessed for the purpose of research. Chinese word segmentation is a very important task in Chinese natural language processing because there

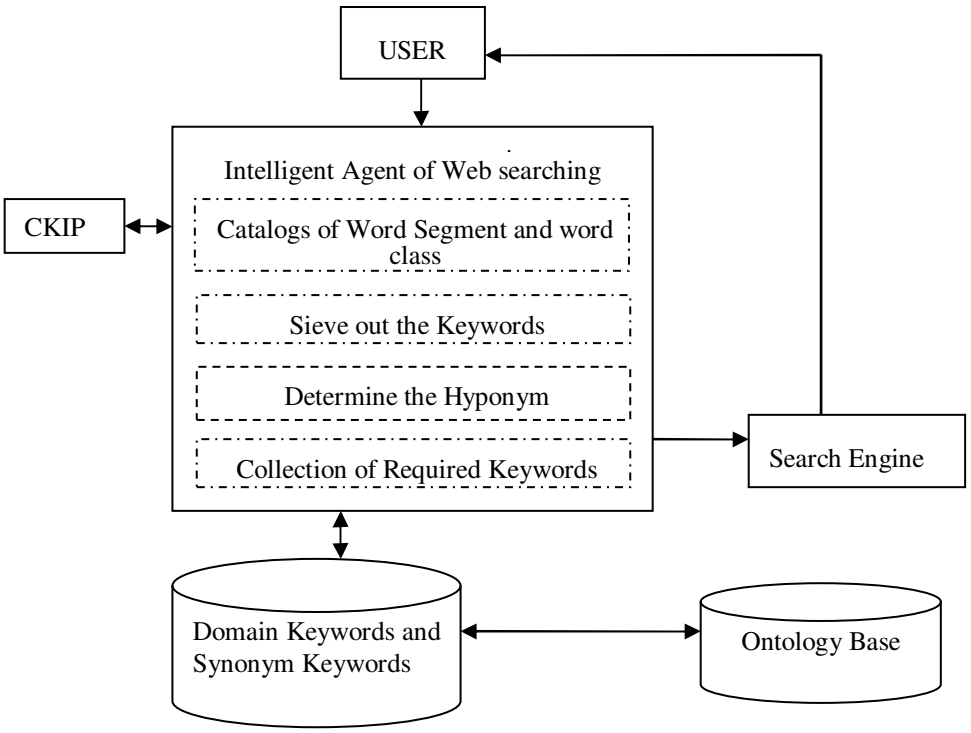

Fig. 2 System Framework

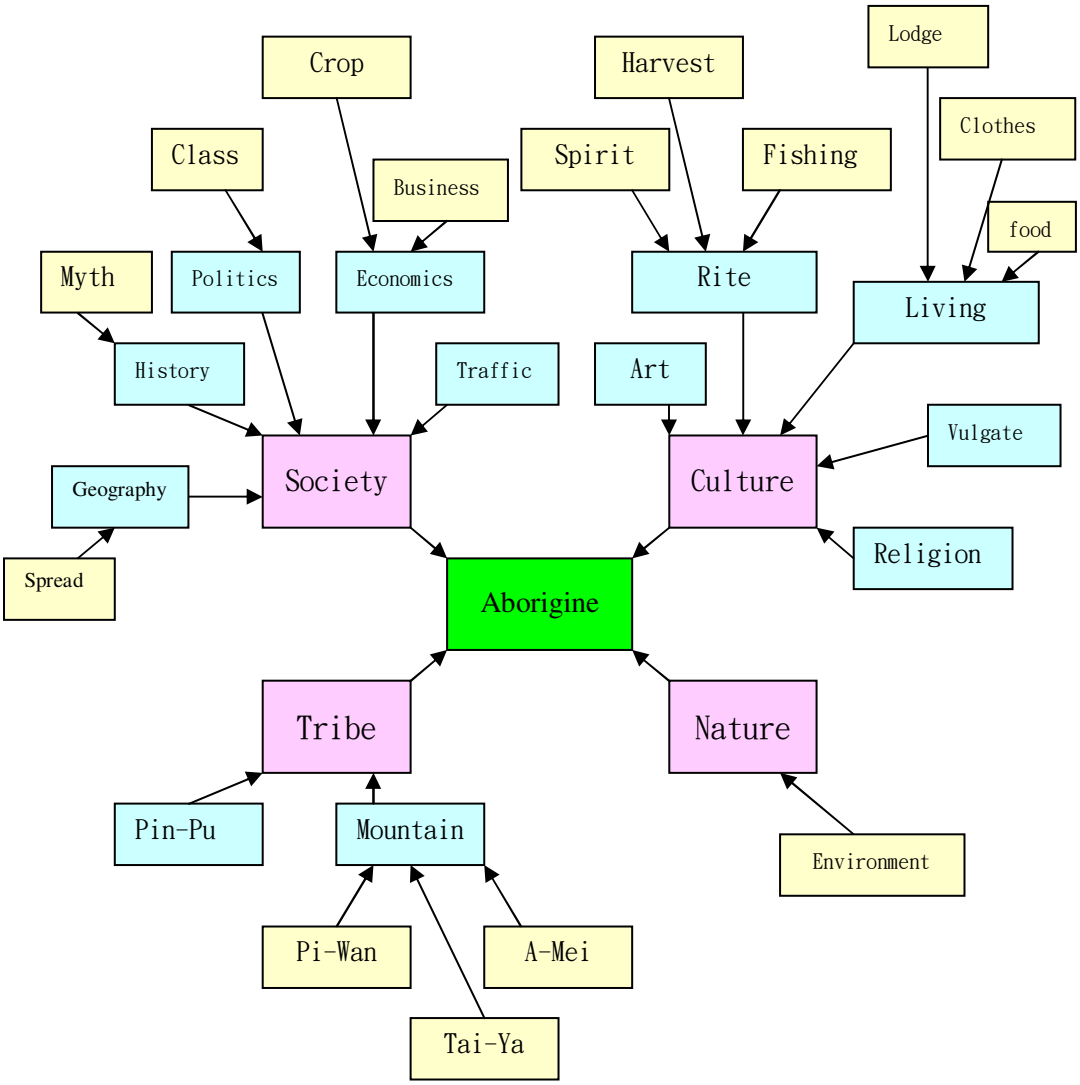

Fig. 3 Ontology for Taiwanese Aborigine culture and history

is no clear separation symbol such as white space between words in a Chinese sentence [7]. The API offered by CKIP not only performs the segmentation of words, it also analyzes the context of each word and determines the part of speech for each word. Only verbs and nouns are of interest since nouns are candidates for keywords and verbs provide additional contextual information to the search engine [5].

3.1 Ontology Base and Keywords Repository

The domain ontology shown in Fig. 3 describes the content of a learning unit, Taiwanese culture and history, in the social study course of the elementary curriculum. The materials come from the Social Studies Materials of 5th Grade of Elementary School, 2006 academic year.

Base on the ontology of Taiwanese culture and history, we created a database of the keywords and defined the keywords of each clerisy, as well as hyponym and synonym.

1. Synonyms are different words of similar meaning. For the purpose of search, our definition of synonyms is not exactly the same as that in a dictionary. For example, to search for "worship ceremony" of the Sai JiaTribe, the word "worship" or "worship little wizards" belong to the same word class. That is, they are synonyms.
2. Hyponyms implied by the ontology structure may be useful in making a search more specific. For example, by inputting keywords such as "Sai Jia Tribe", the hyponyms "distribution", "worship ceremony" and "historical relics" are provided so that users may choose to narrow their search.

After CKIP breaks up a search string, the selected nouns are matched with the vocabulary in the keyword repository to determine keywords and proper hyponyms.

3.2 Intelligent Agent

Fig. 4 illustrates how the Intelligent Agents processes the user input and coordinates the operations of components of the system. After receiving the search string by users, the system invokes the Academia Sinica's CKIP API, passes the search string as a parameter, as shown in Fig. 5. Fig. 6 shows a typical result returned by CKIP with separated words and parts of speech in return. For example, a string "I want to find aborigines" will be broken into "I", "want to", "find "and "aborigines."

After breaking up the string, only nouns and verbs are retained, and all other parts of speech such as pronoun, preposition "of" are deleted. Simple verbs that consist of a single Chinese character are deleted as well since they convey no or little specific meanings. Only nouns and meaningful verbs are saved. Take the above-mentioned example, only the word "aborigines" remains since "want to" is a simple verb in Chinese. Moreover, by comparing with the vocabulary in the keyword repository, only nouns that are defined in the repository are selected as keywords.

Base on the keywords chosen by users, the system will list its hyponym in order to let user add the hyponym into the set of keywords if he/she sees proper. After selecting any hyponym, the system will determine whether there is further hyponym or not. If there is, another hyponym will be provided for students to choose. If not, all of the keywords, which are chosen before, will all be sent to the search engine for searching. During the process of choosing keywords, users can decide whether to continue or not. Fig. 7 is the Process of using hyponym for searching. The collected keywords are sent to Google through Google API.

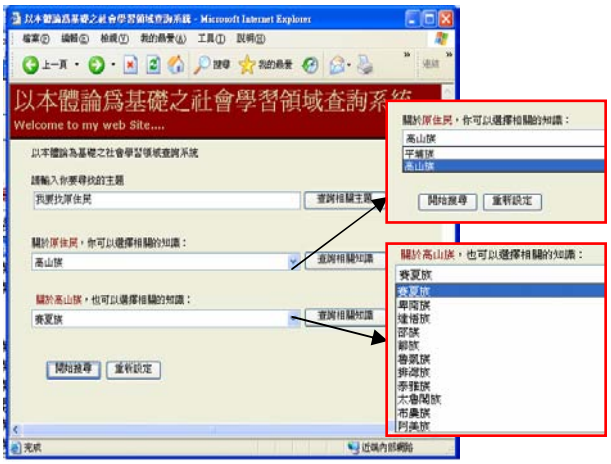

Fig. 7 Using Hyponyms

4 Experimental Results

We conduct five simple experiments to computer our prototype with the traditional Goggle search. The experimental subjects are fifty nine students in the fifth Grade. They are divided into two groups. One group of student uses our prototype and the other group uses traditional Google.

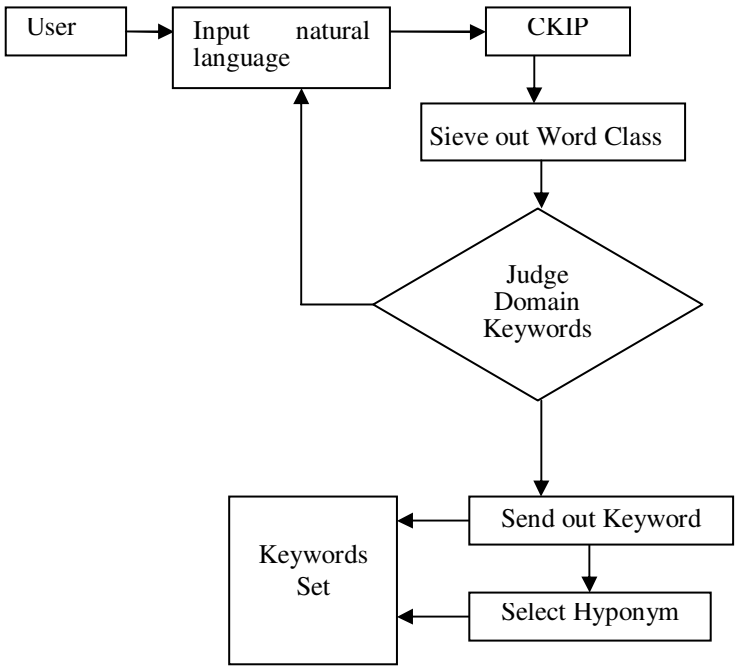

Fig. 4 Flow Chart of Intelligent Agents

```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<option showcategory="1" />
<authentication username="iis" password="iis" />
<text>我要找原住民的祭典</text>
</wordsegmentation>
```

Fig. 5 CKIP API Format

```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<processstatus code="0">Success</processstatus>
<result>
  <sentence>我(Nh)  要(D)  找(VC)  原住民(Na)  的(DE)  祭典
(Na)  ，(COMMACATEGORY) </sentence>
</result>
</wordsegmentation>
```

Fig. 6 Results returned by CKIP

The first experiment is to find one to two worship ceremonies of the aborigines. The question is "What are the worship ceremonies of the aborigines?" The search strings inputted by the students varies somewhat. However, none of them use keywords and redundant words such as "of", "is", or "find" are often found in their search strings. On the other hand, some copy the entire question as the search string. If the entire question is inputted as the search string, our system would resolve keywords "aborigines" and "worship ceremonies". The question for another experiment is "Where are historic spots built by Dutchman in Taiwan?" The keywords found by our system are "Dutchman" and "Historic spot."

We only consider the first twenty items returned by the search engines since these students cannot study the information on the screen for a long time because they are immature in mental and patience. The search results are then examined by a teacher to judge whether a page contains the desired information, that is, a hit. A t test is then applied to examine the performance of two groups. Tables 1 through 5 show the results of the experiments. Four results show that our ontology-based prototype performs significantly better than the tradition search.

5 Conclusion

This work addresses the application of ontology concept to improve searching mechanism that may benefit students in elementary school or even junior high school. Proper keywords derived from ontology can be derived from search strings in natural language which is a common practice for children in these ages.

Our prototype consists of an intelligent agent that transforms a search string into verbs and nouns, sieves out keywords and hyponyms in the repository built from ontology. From the result of two simple experiments, this system could render search results that contain significantly more hits in the first twenty items in a specific domain. This makes it an ideal search mechanism for students exploring information and knowledge that are related to the learning unit of interest. Furthermore, through the guidance of the inquiry interface of this system, students may gradually realize the notion and importance of "keywords", which would gradually improve their searching skills. In the future, this study will focus on extending the application of the proposed mechanism to cover a more general domain.

References:

[1] B. Chandrasekaran, J.R. Josephson, V. Benjamins, What Are Ontologyies, and Why Do We Need Them? *IEEE Intelligent Systems*, May-June, 1999, pp. 20-26.

[2] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition*, vol. 5, issue 2, 1993, pp. 199-220.

[3] N. Guarino, Formal Ontology and Information Systems, in *Processing of the 1st International Conference on Ontology-driven information systems*, 1998, pp. 3-15, Trento, Italy.

[4] C.S. Lee, J.X. Liao, Y.H. Kuo, A semantic-based concept clustering. mechanism for Chinese news ontology construction, In *proceedings of the International Computer Symposium*, 2002, Taiwan.

[5] C.H. Lin, The study of Word Segmentation of Chinese based on Hidden Markov Model, *Master's thesis*, Department of Information Engineering, National Central University, 2006, Taiwan.

[6] Y.L. Liu, An action research on how students in elementary school search and organize information from WWW, *Master's thesis*, Graduate Institute of Educational Technology, National Chia-Yi University, 2002, Taiwan.

[7] W.Y. Ma, K.J. Chen, A bottom-up Merging Algorithm for Chinese Unknown Word Extraction, in *Proceedings of ACL workshop on Chinese Language Processing*, 2003, pages 31-38.

[8] J. Schacter, K. Gregory, W.K. Chung, A. Dorr, Children's Internet searching on complex problem: performance and process analysis. *Journal of the American Society for Information Science*, Vol. 49, No. 9, 1998, pp. 840-849.

[9] A. Spink, D. Wolfram, B.J. Jansen, T. Saracevic, Searching the web: The public and their queries. *Journey of the American Society for Information Science*, 52(3), 2001, pp. 226-234.

[10] M. Uschold, M. Gruninger, Ontologies: Principles, methods and applications, *The Knowledge Engineering Review*, Vol. 11, No. 2, 1996, pp. 93-136.

Table 1. Result for "the worship ceremonies of the aborigines"

group	**Student number**	**Average hit**	**Standard error**	*t*	*p*
Traditional	29	3.69	2.714	6.829	.000***
Ontology-based	30	8.43	2.622		

*** $p < .001$

Table 2. Result for "the cultural characteristics of the aborigines"

group	**Student number**	**Average hit**	**Standard error**	*t*	*p*
Traditional	29	2.03	1.569	7.887	.000***
Ontology-based	30	8.10	3.898		

*** $p < .001$

Table 3. Result for "the historic spots built by Dutchman in Taiwan"

group	**Student number**	**Average hit**	**Standard error**	*t*	*p*
Traditional	29	3.55	2.515	10.122	.000***
Ontology-based	30	10.93	3.051		

*** $p < .001$

Table 4. Result for "the historic spots built by Dutchman in Taiwan"

group	**Student number**	**Average hit**	**Standard error**	*t*	*p*
Traditional	29	3.10	3.086	2.242	.029*
Ontology-based	30	5.27	4.218		

* $p < .05$

Table 5. Result for "the historic spots built by Dutchman in Taiwan"

group	**Student number**	**Average hit**	**Standard error**	*t*	*p*
Traditional	29	5.38	5.525	1.167	.249
Ontology-based	30	6.83	3.869		