Implementation of an Ontology-based Information Retrieval Model in the Classroom Setting

Sook-Young Choi
Woosuk University
Samrae-eup, Jeonbuk, Republic of Korea
sychoi@woosuk.ac.kr

Abstract: This study used ontology as domain knowledge for generating search terms and extending queries in order to improve students' information-seeking performance. It analyzed whether the ontology-based information searching model affects students' effectively searching for meaningful information to carry out their project in the classroom setting. Seventy students from the sixth grade of a primary school located in Jeonju, Korea participated in the experiment for the main study. The experiment results illustrated that the amount of relevant information sought by the ontology-based information retrieval method was significantly greater than that of the existing information retrieval method. The students of the ontology-based information retrieval group reported that the ontology-based information retrieval model helped them to search out information more effectively.

Introduction

The Web has provided students with new opportunities for obtaining information around the world (Bilal, 2000). Especially, with project-based approaches, information resources, like the Web, can play a critical role in the learning environment. As the Web has been used as a major source of information in many educational settings, it may be required to help students find meaningful Internet resources and make sense of what they have discovered (Lee, 2004).

However, the current technology which supports keyword-based search techniques in information retrieval, machines cannot understand and interpret the meaning of the information, which is how most Web information is represented today (Noy et al., 2001). Documents are retrieved if they contain keywords specified by the user. However, many documents contain the desired semantic information even though they do not contain the user-specified keywords (Khan, 2004). Accordingly, users require a Semantic Web to express information in a precise and machine-interpretable form ready for software agents to understand and to process what the terms describing the data mean and thus to help users find practical information and use it.

The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation (Berners-Lee, 1999). How can one make the Semantic Web possible in a way that computer could understand the semantic meaning of the information presented on the Web? Ontology plays a pivotal role here by providing a source of shared and precisely defined terms that can be understood and processed by machines (Bechhofer et al., 2001).

A typical ontology consists of a hierarchical description of important concepts and their relations in a domain (Noy et al., 2001). The ontology captures one or more experts' conceptual representation of a domain expressed in terms of concepts and the relationships among them. Ontologies are important because they provide a common and explicit framework for sharing and using knowledge (Chung et al., 2003). Ontologies were first developed as a part of the artificial intelligence research efforts to facilitate knowledge sharing and reuse. The use of ontologies has extended recently to fields such as information retrieval, knowledge management, and e-commerce (Chung et al., 2003).

Statement of the Problem

Students get the information needed for project-based learning through Web searching in a classroom setting. However, most information searched and located on the Web results in irrelevant data that are not specifically useful with the original intent or purpose (Large & Beheshti, 2000). Moreover, the ever-increasing amount of information being created on the Web requires better searching techniques for effective search and retrieval (Khan et al., 2004). A key problem is that students can be easily overwhelmed by the amount of information available via electronic means. The transfer of irrelevant information in the form of documents retrieved by an information retrieval system and that is of no use to the students, wastes network bandwidth, and may frustrate students.

This condition is a result of inaccuracies in the representation of the documents in the Web as well as confusion and imprecision in students queries because students are frequently unable to express their needs efficiently and accurately (Schacter et al., 1998; Bilal, 2000; Bilal, 2001). These factors contribute to the loss of information and to the provision of irrelevant information. Lee (2004) discussed the shortcomings of Web search engines and Web browsers as learning environments and provided an experimental Web search environment intended for helping students to use Web-based information efficiently by employing visualization and data mining techniques.

Although elementary school students can procedurally operate an online retrieval system, research has shown that they are lacking in conceptually-driven information-seeking skill. Previous studies have shown that children rarely employ systematic search strategies and spend little time on planning their search (Schacter et al., 1998; Large et al., 1999; Large & Beheshti, 2000).

Several researchers have emphasized the importance of user training and specially designed user centered interfaces if children are to exploit Web-based information resources more effectively (Bilal, 1999; Hirsh, 1999). If an interface for informational retrieval, in which students can see the domain knowledge and choose the appropriate terms for searching, is provided for students when they need to search the information for their project, they would accomplish more effective information retrieval. Accordingly, they could perform their project more successfully.

The importance of prior domain knowledge for learning has been well documented (Alexander & Judy, 1988; Greene, 1995). Learning seems to be both easier and more successful when students access relevant knowledge of the domain (subject matter) that can be used for meaningful integration of new information (Brown et al., 1983; Greene, 1995). Domain knowledge represents a familiarity with existing subject matter and experiences related to the domain in which one is searching or learning (Hill & Hannafin, 1997). Domain knowledge is required for learners to generate search terms that are relevant to their specific project topic.

However, ontology is defined as a formal, semantical specification of a conceptualization of a domain of interest. Ontologies are used to describe the semantics of information exchange and are meant to provide an understanding of the static domain knowledge (Luca & Nurnberger, 2004). Therefore, ontology can be used to describe a domain knowledge required for students' generating search terms corresponding to their specific project topic. In several studies, ontologies have been used to improve retrieval performance by semantically expanding queries (Vallet et al. 2005; Khan et al., 2004; Varga et al, 2003). Nevertheless, there seems to have been no studies that applied ontology to learning environment in order to help students find meaningful Web source and effectively perform their project.

Purpose of this Study and Research Questions

The purpose of this study was to analyze the effect of the ontology-based information retrieval model as a learning supplementary tool. The ontology-based information retrieval model was used for students to search information on the Web for performing their projects in the classroom setting. The use of ontology-based information retrieval model was intended to improve the students' information-seeking performance more effectively than the existing information searching model and to improve project-based learning.

For analyzing the effect of the system, first, this study constructed an ontology in which learning concepts for a subject and relations among them are defined and organized, developing an interface in which Web browsing using the ontology can be conducted. Then, this study applied the interface to information-searching for project-based

learning in the classroom setting. In order to analyze the effectiveness of the ontology-based information retrieval model for students' information-seeking, this study compared the searching performance of the model with the one using the existing Web searching method. Additionally, in order to explore its effect as a learning supplementary tool for project-based learning, this study investigated whether the ontology-based information system could be helpful for students to perform their projects.

The purpose of the study was to answer the following questions:

1. To what extent is there a significant difference between the amount of relevant information sought by the ontology-based information retrieval and the one by the existing information retrieval?
2. To what extent is there a significant difference between the relevance rate of the bookmarked documents (i.e., saved document for their project) sought by ontology-based information retrieval and the one by the existing information retrieval method?
3. How do students perceive the usefulness of an ontology-based information retrieval model in searching information for their project?
4. What is the relationship between information-searching using the ontology-based information model and students' project products?

Method

This study used ontology as a domain knowledge for generating search terms and extending queries in order to improve students' information seeking performance. Accordingly, it analyzed whether the Ontology-based Information Retrieval (OIR) model affects students in effectively searching for meaningful information to carry out their project.

This study carried out with a mixed research design approach. In order to examine if there was a difference between the result sought by the Ontology-based Information Retrieval (OIR) method and the one by the Existing Information Retrieval (EIR) method, a research design with a post-test of both a control group and an experimental group was used. In addition, in order to explore if the ontology-based information retrieval model was helpful for students to perform their project-based learning, interviews with participating students was conducted. Furthermore, before the main study, a pilot study was conducted to assess whether the OIR model was workable and to refine the data collection procedure and instruments of this study.

Participants

Eight students from the sixth grade of a primary school located in Jeonju, Korea participated in a pilot study and seventy students of two classes from the sixty grade of another primary school located in Jeonju, Korea participated in the main study. These students were chosen as subjects in the experiment because the social science course of the sixth grade has chapters that are appropriate to Web searching for project-based learning.

Instruments

This study chose a chapter from a text of sixth grade social science course as a target domain and constructed ontologies for the learning concepts of the chapter with the help of teachers. The topic of the project that was provided for the Web search was selected after discussions between the researchers and the teachers.

The experimental post-test research design was used to analyze the retrieved information. The experimental group was the students who used the OIR method and the control group was students who used the EIR method. The independent variable was Web searching using the OIR method. The dependent variable was the amount of the relevant information in the first 30 hits of every search, the pertinence rating of the bookmarked documents sought, and students' perception toward the information searching by each information retrieval method.

Precision and recall are the basic and traditional measures used in evaluating search strategies (Gordon, 1999). Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant

records retrieved. However, measuring recall is nearly impossible in Web search because it is impossible to know how many relevant records exist in the Web (Shafi & Rather, 2005). Thus, this study employed only precision to evaluate the amount of the relevant information sought by each search method. The first 30 hits of every search, the number of documents that users are reasonably willing to look at after search (Hersh, 2000), were assessed for the presence of the gold standard answer (predefined right answer) to the question.

The gold standard, where all documents are judged as relevant or irrelevant to each query, was constructed manually by experts to use this measure of precision. A result page was considered to contain the standard answer if the answer could be found like that no more than one link from the initial page and at least 90% of the established gold standard answer was present. This criterion was derived from the one suggested by Plovnick & Zeng (2004).The total number of assessed hits containing the standard answer was recorded and the fraction of the assessed hits containing the gold standard answer was calculated.

A relevance rating measure was developed for both students and observers to score pertinence of the bookmarked document sought by each search method and was adapted from the relevance scale Schacter et al. (2000) developed. As the pertinence rating measure, some items such as usefulness, accuracy, depth, and relevance were considered. Students and observers were asked to rate each bookmarked document on a five-point scale. There were ten-minute interviews with ten selected students in the OIR group after the completion of the project to examine how the students perceive the OIR model in searching information for their project.

Procedure

Students of the experimental group attended a 20-minute training session in which sample Web searches using the OIR interface was presented. This experiment was conducted in a single session. The session lasted one class period (approximately 40 minutes). While each student conducted his or her information search, two observers observed and recorded the student's search behavior and noted the identity and score the pertinence rating of the bookmarked result. If there was no result in searching, the pertinence rating measure and the top 30 hits precision were assigned a convenience score of zero in this study. Additionally, students recorded their keywords used for searching and made bookmarks on documents related to project topic. Then, they were requested not to exceed five book marks. Students and observers were asked to rate each bookmarked document on a five-point scale. The amount of the relevant information sought (precision) was calculated with the first 30 hits of every search by experts after the session. The search result was sought based on keywords students recode.

The document for the project was submitted at the end of project and was graded by the teacher. After the completion of the project, the students who submitted a completed project were given a questionnaire to survey their opinions on the Web search conducted in the experiment. Then, they were interviewed with open-ended questions, and the interviews were recorded anecdotally and subsequently transcribed. The questions were designed to gather data about the students' attitude to the ontology-based information retrieval method.

For reliability, the internal consistency of the pertinence rating measure of the bookmarked document and the questionnaire to survey students' opinion on the Web search conducted in the experiment were accessed for both the experimental and control groups. Cronbach's alpha, which is the most common reliability estimate, was used to calculate the inter-item reliabilities of the relevance rating measure and questionnaire. High inter-item consistency of the relevance rating measure was obtained (Cronbach's alpha = .82) on students' bookmarks sought by each search method. The questionnaire consists of Part I for checking students' information searching skill and Part II for checking students' perception toward the Web search conducted in the experiment. High inter-item consistencies of Part I (Cronbach's alpha = .84) and Part II (Cronbach's alpha = .81) were obtained. The inter-rater reliabilities of the scores between two experts measuring the relevance rate of each student's bookmarks were assessed. The inter-rater consistency of the experts scores ranged from .83 -.92.

For face validity, a sample group which consists of five teachers and five students was asked to review the initial relevance rating measure and the questionnaire. Then, for content validity, experts in the field of information retrieval were contacted and asked to review the initial relevance rating measure and the questionnaire to indicate if the items need to be modified or deleted. Additionally, ontologies for the learning concepts were reviewed by experts and teachers in the field of social science.

Results and Implication

Research Question 1

In order to answer the first research question, students conducted an information search for the project topic and then the amount of relevant information sought was calculated with the first 30 hits of every search. The independent sample t-test shown in Table 1 indicates that for 35 pairs of students, the mean score (M=13.19, SD=4.22) of the OIR group was significantly greater than that (M=10.41, SD=4.30) of the EIR group (t=2.717. p<.01). There was statistically a significant difference between the amount of relevant information sought by the OIR model and that of the EIR model. In other words, the amount of relevant information sought by the OIR model was more than that of the EIR model.

The statistical analysis of the mean difference between the amount of relevant information sought by the OIR method and that of the EIR method indicated that using the OIR method was more effective than using the EIR one in information searching. The possible explanation for greater amount of relevant information retrieved by the ontology-based group could be the availability of appropriate search terms through concepts to the project topic and their relationships shown by the tree-like type in its interface. In other words, the students of the OIR group could effectively generate appropriate search terms to seek information for their project through the interface and thus find more relevant information than the EIR group. This feature may be related to reduction of irrelevant information to the project topic and to achievement of better precision and recall in information retrieval.

	OIR group (N=35)		EIR group (N=35)		T	Sig.
	M	SD	M	SD		
Amount of relevant information sought (precision)	13.19	4.22	10.41	4.30	2.717	.008**

** p<.01

Table 1: Amount mean of relevant information sought by the OIR group and the EIR group

Research Question 2

In order to answer the second research question, students conducted an information search for the project topic, making five bookmarks on documents related to it, and rating them with a five-point relevance scale. Observers rated each bookmarked document with the pertinence scale. The independent sample t-test on Table 2 shows that the mean of expert rated pertinence score of the OIR group (M=3.57, SD=.32) was significantly greater than that (M=3.26, SD=.54) by the EIR group (t=2.909. p<.01). There was statistically a significant difference between the expert rated pertinence score of the bookmarked documents sought by the OIR model and that of the EIR model.

The statistical analysis of the mean difference in the relevance scores of the bookmarked documents between the OIR group and the EIR group also indicated that the students of the OIR group could find more relevant information related to the project topic than the ones of the EIR group.

This result indicates that the OIR model helped students to effectively search information. That is, the students of the OIR group could more easily generate appropriate search terms to find information for the project through the interface of the model. Accordingly, they could find more information which was relevant and useful for conducting the project than the students of the EIR group. In addition, according to the analysis of the interview, the OIR group tended to consider that they were likely to find information more quickly. This may be because the information provided in the interface of the OIR model, which shows important concepts and their relationships, could reduce the amount of time for students to generate appropriate search terms.

	OIR group (N=35)		EIR group (N=35)		t	Sig,
	M	SD	M	SD		
Relevance score	3.57	.32	3.26	.54	2.909	.005**

** p<.01

Table 2: Mean of expert rated relevance score of the bookmarked documents sought by the OIR group and the EIR group

Research Question 3

In order to answer the third research question, the survey, which contained five questions related to the students' perception of each information retrieval model, was administered to students in both groups after the completion of the project. Table 3 shows the result of independent sample t-test performed on each item from the survey. According to Table 3, there was a significant difference in each mean score of five questions between the two groups. The independent sample t-test concerning students' perception toward the amount of relevant information sought shows that the mean score (M=3.55, SD=.54) of the OIR group was significantly greater than the one (M=3.10, SD=.59) of the EIR group (t=3.270. p<.01).

The statistical analysis of the mean difference in the students' perception as to each information searching model indicated that more students of the OIR group positively perceived the results of the information searching and their information searching method than the students of the existing group. In addition, the interview results backed up these results. The students in the OIR group generally perceived that using the OIR method helped them to make search terms to find information and thus it was useful in searching for information for their project. In addition, four out of the ten students perceived that the OIR model was helpful for them to understand the subordinate concepts and their relationships for an important learning concept since the information about the project topic and its related concepts were shown by the tree-like diagram in the interface of the model.

	OIR group (N=35)		EIR group (N=35)		t	Sig.
	M	SD	M	SD		
Students' perception of each information search	3.55	.54	3.10	.59	3.270	.002**

* p<.05, ** p<.01

Table 3: Mean of students' perception toward each information search of the ontology-based information retrieval group and the existing information retrieval

Research Question 4

In order to discover the relationship between information searching using the OIR model and the project product, the independent sample t-test was performed on the project grade score for each group. The independent sample t-test on Table 4 shows that the mean score of project grade (M=7.09. SD=1.52) of the students of the OIR group was greater than the one (M=4.77. SD=2.27) of the students of the EIR group (t=5.000, p<001). There was statistically a significant difference between the project grade of the OIR group and that of the EIR group.

The statistical analysis shows that the project grades of the students of the OIR group were greater than those of the EIR group. The statistical analysis of the mean difference in the students' project grade between the OIR group and the EIR group indicates that the OIR model helped the students to effectively find information for their project more than the EIR model. The possible explanation for this result may be that the OIR model helped the students to

generate appropriate search terms and thus the students could find more information which was relevant and useful for conducting the project than those of the EIR group. Accordingly, the students of the OIR group could probably perform the project better than those of the EIR group. In addition, the students in the OIR group could probably find the information for the project more quickly than those in the EIR group and thus they could concentrate on drawing up the report for the project more than those in the EIR group.

	OIR (N=35)		EIR (N=35)		t	Sig.
	M	SD	M	SD		
Project grade	7.09	1.52	4.77	2.27	5.000	.000***

***p<.001

Table 4: Mean of project grade of the OIR group and the EIR group

Conclusion

Although the Web has been used as a major source of information in many educational settings, it is hard for students to use the data for their project because most information searched and located on the Web result in irrelevant data that are not specifically useful to the original intent or purpose (Large & Beheshti, 2000). Therefore, information retrieval models have been required that could help students more effectively search information on the Web. In this article, an ontology-based information retrieval (OIR) model, which was intended to improve students' information-seeking performance more effectively than the existing information retrieval (EIR) model, was applied to information searching for project-based learning in the classroom setting and its effectiveness was analyzed. The analysis result showed that the ontology-based information retrieval model helped students to effectively find information and thus better perform their project. This result indicates that the ontology-based information searching model could be used as a learning supplementary tool for project-based learning.

References

Alexander, P.A., & Judy, J. (1988).The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58, 375-404.

Basili, C. (1995). Subject searching for information :What does it mean in today's Internet environment? *The Electronic Library*, 13, 459-466.

Bechhofer, S., Horrocks, I., Goble, C., & Strvens, R. (2001). OilEd : A reasonable ontology editor for the Semantic Web. *Proceedings of 14 International Workshop on Description logics.* Stanford, USA.

Berners-Lee, T. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.* San Francisco: HarperCollins.

Bilal, D. (1999). Web search engines for children: A comparative study and performance evaluation of Yahooligans!, ask Jeeves for Kids and Super Snooper, *Proceedings of the 62nd ASIS Annual Meeting,* USA.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine : I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7), 646-665.

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine : II. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 52(2), 118-

Chung, G. K., Niemi, D., & Bewley, W. (2003). Assessment applications of ontologies, *Paper presented at the annual meeting of the American Educational Research Association,* Chicago, IL.

Greene, B.A. (1995). Comprehension of text in an unfamiliar domain: Effects of instruction that provides either domain or strategy knowledge. *Contemporary Educational Psychology,* 20, 313-319.

Gordon, M. & Pathak, P. (1999). Finding information on the World Wide Web: The retrieval effectiveness of search engines. *Information Processing and Management,* 35, 141-180.

Hill, J.R., & Hannafin, M.J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational Technology Research and Development,* 45(4), 37-64.

Hirsh, S.G. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, 50(14), 1265-1283.

Khan, L., McLeod, D., & Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The VLDB Journal*, 13, 71-85.

Large, A., Beheshti, J., & Breuleux, A. (1999). Information seeking in a multimedia environment by primary school students. *Library and Information Science Research,* 20, 343-376.

Large, A., & Beheshti, J. (2000). The Web as a classroom resource : Attitudes from the users. *Journal of the American Society for Information Science*, 51(12), 1069-1080.

Lee, Y. (2004). Concept mapping your web searches: A design rationale and Web-enabled application. *Journal of Computer Assisted Learning*, 20, 103-113.

Luca, E. W. & Nürnberger, A. (2004). Ontology-Based Semantic Online Classification of Documents: Supporting Users in Searching the Web. *Proceedings of the European Symposium on Intelligent Technologies.*

Noy, N. F., Sintek, M., Decker, S., Crubezy, M., Fergerson, R.W., & Musen, M. (2001). Creating Semantic Web contents with Protégé-2000*, IEEE Intelligent System*, 60-71.

Plovnick, R.M. & Zeng, Q. T. (2004). Reformulation of Consumer Health Queries with Professional Terminology: A Pilot Study. *Journal of Medical Internet Research, 6,3.*

Schachter, J., Chung, G. K., & Dorr, A. (1998). Children's Internet searching on complex problems: Performance and process analyses. *Journal of the American Society for Information Science,* 49(9), 840-849.

Suchard , M., Adamson, S., & Kennedy, S. (1997). Netpoints: Piloting patient attitudinal surveys on the Web. *BMJ* 315:529. Retrieved January 27, 2006 from http://bmj.bmjjournals.com/cgi/content/full/316/7124/72

Vallet, D. Fernandez, M. & Castells, P. (2005). An Ontology-based information retrieval model. *Lecture Notes in Computer Science*, 45.

Varga, P., Meszaros, T., Dezsenyi, C., & Dobrowiecki, T.P. (2003). An Ontology-based information retrieval system. *Lecture Note in Computer Science.*