An Analysis of the Effect of an Ontology-Based Information Searching Model as a Supplementary Learning Tool

Sook-Young Choi†

ABSTRACT

This study analyzed whether the ontology-based information-searching model affected the ability of students to effectively search for meaningful information to carry out their projects. The experiment results illustrated that the amount of relevant information sought by the ontology-based information retrieval (OIR) method was significantly greater than that of the existing information retrieval (EIR) method. In addition, the relevance rate of the bookmarked documents sought by the OIR method was significantly greater than that of the EIR method. Interviews showed that the OIR model was helpful for students to effectively find information and thus, it helped them to complete the project more easily. Furthermore, the OIR model was beneficial for them to understand the subordinate concepts and their relationships for an important learning concept. The results of this study indicate that the OIR model could be used as a supplementary learning tool for project-based learning.

Keywords: Ontology-based information-searching, Amount of relevant information, Supplementary learning tool

학습 보조 도구로서 온톨로지 검색 모델의 효과 분석

최숙영†

요 약

본 연구에서는 온톨로지기반의 정보검색 모델이 프로젝트를 수행하기 위해 필요한 정보를 학생들이 검색하는데 미치는 효과를 분석하였다. 본 연구의 분석 결과는 온톨로지 기반의 검색에 의해 검색된 관련 정보의 양이 기존의 검색에 의한 것보다 더 많음을 볼 수 있었다. 뿐만 아니라, 온톨로지 기반의 검색에 의해 찾아진 뷰 뷰 검색에 의해 원재기록 수집 정보의 관련성 기준이 기존의 검색 방법에 비해 높음을 알 수 있었다. 학생들의 인터뷰 결과에 의하면 온톨로지 기반의 검색 모델은 프로젝트 수행을 위한 자료를 검색하는 데 매우 유용했으며, 이에 따라 그들이 프로젝트를 보다 쉽게 수행할 수 있도록 도움이 되었음을 알 수 있었다. 또한, 온톨로지기반 검색 모델에서 제공하는 학습 개념간의 관련성 정보는 검색뿐만 아니라 학습 주제에 관련된 학습 개념들을 이해하는데도 도움이 되었음을 알 수 있었다.

주제어: 온톨로지기반 정보검색 모델, 관련 정보 양, 학습 보조 도구

† 종신회원: 우석대학교 컴퓨터교육과 교수
논문접수: 2011년 01월 03일, 심사완료: 2011년 01월 11일
* 본 논문은 2010년 우석대학교 교내 학술연구비 지원에 의하여 수행되었음.
1. Introduction

The Web has provided students with new opportunities for obtaining information around the world[7]. Especially, with project-based approaches, information resources, like the Web, can play a critical role in the learning environment. 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 [21]. 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 probably caused by two aspects. First, with 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[26]. 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[19]. 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[5]. Ontology plays a pivotal role by providing a source of shared and precisely defined terms that can be understood and processed by machines[4]. A typical ontology consists of a hierarchical description of important concepts and their relations in a domain[26].

Second, since students are frequently unable to express their needs efficiently and accurately[7][31], there are confusion and imprecision in students queries. These factors contribute to the loss of information and to the provision of irrelevant information. 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 [17]. However, 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.

Domain knowledge is required for learners to generate search terms that are relevant to their specific project topic. Here, ontology can be used to describe a domain knowledge required for students’ generating search terms corresponding to their specific project topic, because it is used to describe the semantics of information exchange and are meant to provide an understanding of the static domain knowledge[23]. In several studies, ontologies have been used to improve retrieval performance by semantically expanding queries[32][19]. 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.

2. Literature Review

2.1 The Web and Instruction

The Web has great potential for constructivist learning. As each learner has access to rich resources of information on the Web and processes it, knowledge is constructed and reconstructed[18].

Several studies have examined students’ use of the Web. Roy, Taylor, and Chi(2003) examined how students search for, browse, and learn specific information when using the Internet versus the library. The study concluded that the Internet was superior to
the library in supporting students’ searches for target-specific information. Gunn and Hepburn(2003) studied 12th-grade students’ information-seeking strategies using the Internet as an information source. The survey result showed that most students used few strategies and techniques associated with effective Internet searching. Typically, the students relied on keyword searching. Hansen, Derry, Resnick, and Richardson (2003) investigated how adolescents search for information using the Internet. This study found that adolescents often chose search strings that were either too general or contained misspellings so that they did not always find useful sites.

2.2 Information retrieval

The success of information retrieval is ultimately assessed by the degree of user satisfaction with the quality of the retrieval information in solving the problems[33]. In a user-centered paradigm, user satisfaction can be studied through questionnaires, surveys, or direct observations. In a system-centered paradigm, user satisfaction is measured mainly by retrieval effectiveness. Two basic quantitative measures of retrieval effectiveness are widely used to derive other effectiveness: (a) recall and (b) precision[9]. Recall is the proportion of retrieved relevant documents to the total relevant documents, whereas precision is the proportion of retrieved relevant documents to the total retrieved documents. Recall reflects the system’s power of including all possible relevant documents; thus, its calculation requires knowledge of the relevant and nonrelevant hits in the evaluated set of documents. Precision measures the system’s capability of excluding nonrelevant items; thus, it requires knowledge not just of the relevant and retrieved items but also those items not retrieved[8]. However, it appears that there are no proper methods of calculating the absolute recall of search engines, as it is impossible to know the total number of relevant sources in a huge database like the Web.

Relevance is perhaps the most important and controversial concept in the field of information retrieval. A document’s relevance to an information need could be affected by many factors, such as its subject content, novelty, authority, credibility, availability, and so forth[33]. Relevance is the idea that a particular document may be judged to pertain to a particular query. Relevance may be measured in many different ways, including test matching and expert judging[15].

2.3 Semantic Web and Ontology

The Semantic Web, developed by Berners-Lee (1999), is an extension of the current Web, in which information is given well-defined meaning, better enabling people to work with computers. If a computer understands the semantics of a document, it does not just interpret the series of characters that make up that document; it understands the document’s meaning in a given context[3]. The Semantic Web can be considered either as an efficient way to represent data on the Web or as a database that is globally linked in a manner understandable by machines to the content of documents on the Web. Semantic technologies represent meaning using ontologies and provide reasoning through the relationships, rules, logic, and conditions represented in those ontologies [6].

Ontology captures the conceptual representation of one or more experts of a domain expressed in terms of concepts and the relationship among the concepts. Ontology may also refer to an agreement about shared conceptualizations[11].

To provide navigation aids to assist users’ information exploratory tasks on the Web, it is important to provide semantic structure of the information environments. In this aspect, ontology can play an important role. That is why it is useful for structuring concepts with their relationships and providing an understanding of the static domain knowledge that facilitates knowledge sharing and reuse[23]. An attempt has been made to match users’ exploratory tasks by providing an ontology-based model[22]. Ontologies have been employed to achieve better precision and recall in text retrieval systems[10]. Among the ontology-based information retrieval systems, WordNet[25] used query expansion
through the use of semantically related terms with a generic ontology. It was shown to be potentially relevant to enhanced recall as it permits matching a query to relevant documents that do not contain any of the original query terms.

3. Research purpose

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 (OIR) model was used for students to search information on the Web for performing their projects in the classroom setting. The use of the OIR model was intended to improve the students’ information-seeking performance more effectively than the existing information searching (EIR) 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 OIR 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 OIR model and that of the EIR model?
2. To what extent is there a significant difference between the relevance rate of the bookmarked documents (i.e., saved documents for their project) sought by the OIR model and that of the EIR model?
3. How do students perceive the usefulness of an OIR model in searching for information for their project?
4. What is the relationship between information searching using the OIR model and students’ project products?

4. Methodology

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 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 OIR method and the one by the 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.

4.1 Participants

Eight students from the sixth grade of a primary school located in Jeonju, Korea, participated in a pilot study, and 70 students from the sixth 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 had chapters that were appropriate to Web searching for project-based learning.

4.2 Instruments

This study chose a chapter from the text of a
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 researcher and the teachers.

The experimental posttest research design was used to analyze the retrieved information. The experimental group included the students who used the OIR method, and the control group included the 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 relevance rating of the bookmarked documents sought, and students’ perceptions of the information searching by each information retrieval method. In order to measure the amount of the relevant information, 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, which were the number of documents that users were reasonably willing to look at after search[16], 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 gold standard answer if the answer could be found 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 gold 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 the relevance of the bookmarked document retrieved by each search method and was adapted from the relevance scale Schachter et al. (1998) developed. As the relevance rating measure, some items, such as usefulness, accuracy, depth, and pertinence, were considered. Students and observers were asked to rate each bookmarked document on a five-point scale.

4.3 Procedure

Students in the experimental group attended a 20-minute training session, during which sample Web searches using the OIR interface were 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 students’ search behavior, noted the identity, and scored the relevance rating of the bookmarked result. If there was no result in searching, the relevance rating measure and the top 30 hits precision were assigned a convenience score of 0 in this study. Additionally, students recorded their keywords used for searching and made bookmarks on documents related to the project topic. Then they were requested not to exceed seven bookmarks. 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 recorded. The document for the project was submitted at the end of the project and graded by the teacher. After 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’ attitudes toward the OIR method.

5. Results and Discussion

5.1 Quantitative result

Research Question 1

To answer the 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 was performed on the total score on the amount of relevant information retrieved for each pair to discover whether there was a significant difference between the amount of relevant information retrieved by the OIR model and that of the EIR model. Table 1 shows the result of the independent sample t test performed on the total score on the amount of relevant information. That is, 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). Based on the analysis results, a statistically significant difference existed between the amount of relevant information retrieved by the OIR model and that of the EIR model. In other words, the amount of relevant information retrieved by the OIR model was more than that of the EIR model.

| <Table 1> T-Test Comparison of Amount of Relevant Information Retrieved |
|------------------------|---------------------|------|------|-----|
|                       | OIR group           | EIR group   | t    | p   |
|                       | M     | SD   | M     | SD   |
| (precision)           | 13.19 | 4.22 | 10.41 | 4.30 |
|                       | 2.717 | .008 |

Research Question 2

To answer the 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 relevance scale. The independent sample t test was performed on each relevance score rated by experts and students of the information sought to discover whether there was a significant difference between the relevance score of the bookmarked documents (i.e., saved document for their project) by the ontology-based information retrieval model and the one by the existing information retrieval model. Table 2 shows the result of the independent sample t test performed on the each relevance score rated by experts of the bookmarked documents. Based on the analysis results, there was a statistically significant difference between the relevance score rated by experts on the bookmarked documents sought by the OIR model and that of the EIR model.

| <Table 2> T-Test Comparison of Expert–Rated Relevance Scores of the Bookmarked Documents Retrieved |
|------------------------|---------------------|------|------|-----|
|                       | OIR group           | EIR group   | t    | p   |
|                       | M     | SD   | M     | SD   |
| Relevance score       | 3.57  | .32  | 3.26  | .54  |
|                       | 2.909 | .005 |

Table 3 shows the result of the independent sample t test performed on the expert-rated scores on all items of the relevance rating scale of the bookmarked documents.

| <Table 3> T-Test Comparison of Expert–Rated Relevance Scores on All Items of the Relevance Rating Scale of the Bookmarked Documents Retrieved |
|------------------------|---------------------|------|------|-----|
|                       | OIR group           | EIR group   | t    | df  | p   |
|                       | M     | SD   | M     | SD   |
| Accuracy              | 3.88  | .59  | 3.49  | .54  |
|                       | 2.845 | 68   | .006 |
| Pertinence             | 3.83  | .46  | 3.56  | .46  |
|                       | 2.415 | 68   | .018 |
| Usefulness             | 3.81  | .62  | 3.40  | .45  |
|                       | 3.102 | 68   | .003 |
| Depth                 | 3.54  | .59  | 3.25  | .50  |
|                       | 2.209 | 68   | .031 |

Research Question 3

To answer the research question, the survey, which contained five questions related to the students’ perceptions of each information retrieval model, was administered to students in both groups after the completion of the project. An independent sample t test was performed on the average of mean scores of all questions of the survey to discover whether there was a significant difference in students’ perceptions of each information retrieval model between the EIR group and the OIR group. Table 4 shows the results of the independent sample t test performed on the average of mean scores of all items from the survey. In addition, to probe for further explication of students’ perceptions
toward the OIR model in searching out information for their project, interviews were conducted with 10 students from the OIR group after completion of the project. The interview results are described separately in the following “Qualitative Results” section.

**Research Question 4**
To discover the relationship between information searching using the ontology-based information model and the project product, the grade of the project product between the EIR group and the OIR group was compared. The independent sample t test was performed on the project grade score for each group.

Table 5 shows the result of the independent sample t test performed on the project grade. There was a statistically significant difference between the project grade of the OIR group and that of the EIR group.

6. **Discussion and conclusion**

6.1 **Discussion of Research Question 1**
One explanation for the greater amount of relevant information retrieved by the OIR 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 could 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. This finding is consistent with the results reported in prior studies, which indicated that ontologies could be used to improve retrieval performance by semantically expanding queries[19][32].

6.2 **Discussion of Research Question 2**
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 those of the EIR group. In addition, for each item of the relevance rating scale, all scores of bookmarked documents sought by the OIR method were significantly greater than those by the EIR method.
This result indicates that the students in 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 that was relevant and useful for conducting the project than could the students in 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.

Aligned with the results from this study, previous studies have shown that children rarely employ systematic search strategies and spend little time on planning their search[31]. Besides poor planning, child searchers have difficulty formulating effective queries. That is, they have difficulty constructing effective search terms[21][31]. From this point of view, it is speculated that providing the interface for OIR in this study, in which students could see the domain knowledge and choose the appropriate terms for searching, helped them accomplish more effective information retrieval. This finding is consistent with those found in similar studies, which indicated that domain knowledge influenced search success on all types of task and search skills definitively improved with domain knowledge [17][30].

6.3 Discussion of Research Question 3

The statistical analysis of the mean difference in the students’ perceptions of each information retrieval model between the OIR group and the EIR group indicated that more students in the OIR group positively perceived the results of the search and its method than did those of the EIR group. In addition, the interview results corroborated these results. The students in the OIR group generally perceived that using the OIR model helped them to make search terms to find information and, thus, it was useful in searching for information for their project. In addition, 4 of the 10 interviewed students perceived that the OIR model was helpful for them to understand the subordinate concepts and their relationships for an important learning concept because the information about the project topic and its related concepts were shown by the tree-like diagram in the interface of the model. This finding is consistent with those found in prior studies, which indicated that domain knowledge is important for learning[1][10]. Greene(1995) found that learning would be easier and more successful when students access relevant knowledge of the domain.

Both the statistical analysis and the interview results seem to indicate that the OIR model would be useful not only in information searching for a project, but also in studying learning concepts related to a project topic.

6.4 Discussion of Research Question 4

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 did. 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 that was relevant and useful for conducting the project than those of the EIR group. Accordingly, the students of the OIR group could perform the project better than those of the EIR group. In addition, the students in the OIR group could 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 could.

6.5 Implications

The results of this study indicate that the OIR model could be used as a supplementary learning tool for project-based learning because the OIR model helped students generate appropriate search terms and, thus, students could effectively find information that was relevant and useful for conducting the project. In
addition, the OIR model was helpful for students to understand the subordinate concepts and their relationships for an important learning concept because the information about the project topic and its related concepts were shown by the tree-like diagram in the interface for the OIR. Therefore, ontology can be used not only to describe a domain knowledge required for students to generate search terms corresponding to their specific project topic, but also to help students to understand the subordinate concepts and their relationships for an important learning concept. Based on this result, it is inferred that ontology could be used in various aspects for learning. This finding is consistent with those found in several research studies, which have been recently performed with reference to the use of ontology for learning [20][28][34].

References


Sook-Young Choi
(최숙영)

August 1998  B.S. in Computer Science, Choungbuk National University

February 1991  M.S. in Computer Science, Choungbuk National University

February 1996  Ph.D. in Computer Science, Choungbuk National University

December 2008  Ed.D. in Instructional Technology and Distance Education, Nova Southeastern University

March 1966  ~  Professor, Department of Computer Education, Woosuk University

Research Interests: E-learning, Intelligent Tutoring System, Computer Education

E-Mail: sychoi@woosuk.ac.kr