Chapter 6

Case study: Embedding the automated ontology and enhanced measure into an e-learning system

The main idea of this research work is to improve the process of information retrieval. This can be explained better with help of an application. An important area where information retrieval will have a successful contribution is e-learning. This chapter briefly discusses how the documents are retrieved after embedding the automated ontology and also how the information retrieval is enhanced after embedding the proposed ConSim measure into the e-learning system.

6.1 INFORMATION RETRIEVAL FROM E-LEARNING SYSTEMS AFTER EMBEDDING THE AUTOMATED ONTOLOGY

The objective of the proposed approach is to develop an application that will help a student to interact with a predefined learning system, which give solutions and guidance for queries raised by the user. The learning system should be defined clearly according to the proposed approach. In the conventional e-learning systems, there is limited interaction between a student and the tutor who gives instructions to the student without assessing his knowledge. On the other hand, things are a little different for an adaptive e-learning system. In adaptive e-learning systems, the student provides information about his knowledge of the subject to the system and the tutor analyses it and provides the learning materials accordingly.

A document personalization feature is adapted by the proposed approach based on the user knowledge level. The user is treated with an evaluation test according to the domain selected by him. The evaluation test contains objective type questions, the user has to select the correct answer from the drop down list. There will be a fixed number of questions and according to the number of correct answers and wrong answers, the user is given a score. The correct answers symbolize the user’s
knowledge of the specified domain and the incorrect answers symbolize the user’s need for knowledge in the specified domain. The solution is given according to the level of capability of the student.

![Diagram](image)

**Figure 6.1: E-learning system embedded with automated domain ontology**

The existing e-learning systems have difficulties in generating the concepts map. Most of the existing systems use human generated concept maps or semi automatically generated concept maps. The method proposed in this thesis uses an automatically generated ontology to organize the relevant materials. Automating an ontology reduces the human effort and errors. In this system, the student is asked a number of questions about each topic based on his search query. By assessing his answers, the e-learning system generates lessons for him and the results are discussed in section 6.3.
6.2. INFORMATION RETRIEVAL FROM E-LEARNING SYSTEMS AFTER EMBEDDING THE AUTOMATED ONTOLOGY AND ConSim MEASURE

The proposed approach makes use of a document retrieval system based on the user’s knowledge level as shown in Figure 6.2. The main parameter used by the proposed approach for retrieving the documents is the concept similarity measure. The similarity function generates a set of values for each query given by the user by incorporating the similarity values of the properties and representatives. The concept similarity value per se, is a parameter adequate for retrieving the documents from the documents set. However, the proposed approach retrieves the documents in an adaptive manner, i.e. by evaluating the user’s knowledge level and processing the document retrieval accordingly. The user’s knowledge level is assessed through a user evaluation test and the test supplies two different set of responses, the correct responses and wrong responses. The proposed approach applies a threshold to the concept similarity measures based on the correct and wrong responses. The user score is generated based on the responses.

Based on the level, the documents are retrieved with the help of the automated ontology in such a way that the application generates a dialog box that contains the user’s level of knowledge with respect to the domain and the marks scored by the student. The threshold (th) on the concept similarity measure acts as the deciding parameter to retrieve the documents from the documents set. The threshold value is set based on the user score. If the score is between 1 and 3 then he is classified as a beginner (Stage 1). If the score is between 4 -7 then he is classified as intermediate (Stage 2), 8-10 he is classified as an expert (Stage 3). The retrieval of documents depends on the wrong responses, as the wrong responses indicate that the user has less information regarding the specific domain, so he may need more number of documents to get relevant information regarding the domain. The correct responses will trigger the retrieval of more specific and most relevant documents regarding the domain.
Step 1: Select $P_{\text{similarity}}$ and $R_{\text{similarity}}$

Step 2: Assign threshold ($th$) values to $P_{\text{similarity}}$ and $R_{\text{similarity}}$

Step 3: Check if ($P_{\text{similarity}} > th$) and ($R_{\text{similarity}} > th$)

Step 4: Compute $C_{\text{sim}} = \sqrt{\frac{\alpha R_{\text{similarity}}^2 + \beta R_{\text{similarity}}^2}{2}}$

Step 5: Select correct responses (T)

Step 6: Select wrong responses (W)

Step 7: If ($T > W$) $\parallel$ ($W > T$)

    Calculate the user score based on the responses

Step 8: Based on the user score and $C_{\text{sim}}$ measure, retrieve the documents

Figure 6.2: Proposed e-learning system using concept similarity measure

Figure 6.3: Algorithm for document retrieval
The algorithmic procedure in Figure 6.3 shows the retrieval of documents from the document set according to the user knowledge level and the similarity measure. The value $th$ represents the threshold value assigned to $C_{\text{sim}}$ to retrieve documents from the documents set. The value $T$ represents the number of correct responses and $W$ represents the wrong responses from the user evaluation test. If the user is in stage 1 then the threshold assigned is 0.5, in stage 2 the threshold will be 1.0 and stage 3 the threshold will be 1.5. The keywords which have similarity measure above the specified threshold are selected, and the learning materials are retrieved based on that. The e-learning system designed here is based on the information retrieved from the two ontologies. Even though the information is retrieved based on the cross ontology measure, the presentation of the retrieved information should be based on the user’s interest.

6.3 EXPERIMENTAL RESULTS AND ANALYSIS

The following sections include the experimental evaluations of the e-learning system under different conditions. The experiments were conducted on a system running with Intel core i5 processor, 4GB RAM and 500 GB hard disk. The programs were written and tested using Java program under JDK 1.7.0. The detailed experiments and analysis are discussed in the following sections.

6.3.1 PREPARATION OF E-LEARNING DOCUMENTS

In an e-learning system, the content is the most important of all its features. Since the focus in this thesis is on automation, content preparation was not stressed much and is considered as a part of our extension work. The textual documents used for automation were used as e-learning materials. The retrieval is based on the user query and the user score from the user evaluation test. This data is prepared according to the positive answers that are given by the user from the evaluation test. They are retrieved for the user according to the specifications that have been discussed in the above sections.
6.3.2 RESULTS AFTER EMBEDDING THE AUTOMATED ONTOLOGY

Let us have a look at the user interface (UI) of our proposed method. Initially the student places a query through an user interface. He is asked some questions related to that topic in the query as shown in Figure 6.4.

Figure 6.4: UI: Questioning phase

The student has to answer all the questions asked by the application to assess his knowledge level as shown in Figure 6.5. Based on the number of questions answered he is categorised in to one of the three stages (1 – Beginner, 2 – Intermediate, 3 – Expert).

Figure 6.5: UI: Deciding phase
After all the questions are answered, the marks and stage of the user is displayed as shown in Figure 6.5. Based on the level, the documents are retrieved with the help of the automated ontology in such a way that the application generates a dialog box that contains the user’s level of knowledge with respect to the domain and the marks scored by the student.

![UI: Document Retrieval phase](image)

**Figure 6.6: UI: Document Retrieval phase**

On clicking the “OK” button as shown in Figure 6.6, the documents pop up

![Sample Documents presented to the user](image)

**Figure 6.7: Sample Documents presented to the user**
Figure 6.7 shows the output generated by the application after assessing the student’s knowledge. A set of sample of documents are presented to the user by the e-learning system based on the stage of the user.

6.3.3 RESULTS AFTER EMBEDDING THE CONSIM MEASURE

The experimentation of the proposed approach deals with the assessment of the efficiency based on the performance of the comparison of concepts between two ontologies. The two ontologies that are being discussed here are the automated domain ontology and the domain expert’s ontology. Since we are dealing with e-learning scenario, a knowledge base domain based on DBMS and its components are used. There is similarity in some aspects of the ontology obtained from both the methods. The domain expert ontology is more precise when compared to the automated domain ontology. So, through the experiments, we find similarities between these two ontologies to provide a better result in retrieving the correct data to the user. The user query is matched with both the ontologies and results are selected according to the similarity measure defined in section 4.2.

6.3.4 COMPARISON OF RESULTS

The documents retrieved from the e-learning system with and without the concept similarity (ConSim) measure are compared in terms of the retrieved documents. The well known statistical measures for measuring the performance and accuracy of information retrieved are precision and recall. Precision is the fraction of documents retrieved that are relevant to the user’s need and recall is the fraction of documents relevant to the query and are successfully retrieved.

\[
\text{Recall} = \frac{\text{No of relevant documents retrieved}}{\text{Total number of relevant documents in the database}}
\]

\[
\text{Precision} = \frac{\text{No of relevant documents retrieved}}{\text{Total number of links selected for evaluation}}
\]
Based on a particular query, the materials are retrieved from the e-learning system, with and without embedding the enhanced similarity measure. As soon as the user queries a particular topic, a set of questions is posed to him based on that topic. The score attained by the user is 2. Based on this score the learning materials are retrieved. Table 6.1 shows the recall and precision values for the query “Normalisation”. A total number of 26 relevant documents are present in the database. The precision and recall were calculated based on relevance of information extracted from first 5 links, then top 10 links and so on till top 25 links.

Table 6.1: Precision and Recall measures of a single query

<table>
<thead>
<tr>
<th>Learning materials retrieved</th>
<th>Information retrieval (Without ConSim Measure)</th>
<th>Information retrieval (With ConSim Measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Top 5</td>
<td>4/5</td>
<td>4/26</td>
</tr>
<tr>
<td>Top 10</td>
<td>7/10</td>
<td>7/26</td>
</tr>
<tr>
<td>Top 15</td>
<td>10/15</td>
<td>10/26</td>
</tr>
<tr>
<td>Top 20</td>
<td>15/20</td>
<td>15/26</td>
</tr>
<tr>
<td>Top 25</td>
<td>23/25</td>
<td>23/26</td>
</tr>
</tbody>
</table>

Figures 6.8, 6.9 and 6.10 show the comparison of recall and precision measures, with and without embedding the similarity measure into the e-learning application.

Figure 6.8: Comparison of Recall measures
Figure 6.9: Comparison of Precision measures

The recall and precision graphs prove that by embedding the ConSim measure into the e-learning system it also enhances accuracy of search results. A high precision value indicates the quality of information retrieved.
6.4 CONCLUSIONS

The focus of the research was to automate a framework for ontology construction to enhance the information retrieval. The developed system can be applied to any domain to extract ontology and compare two ontologies. A sample e-learning system developed is discussed in Figure 6.1 and 6.2 and the retrieval results were compared. It was proved that by embedding the enhanced measure, the retrieval of documents was more efficient.