Chapter 5

Derivation of an enhanced cross similarity measure

Ontologies have become a key concept for providing highly relevant lessons to the learner. In this chapter, an enhanced cross ontology measure by comparison of the similarity of a concept in more than one ontology is proposed. A cross ontology is defined as one which consists of the automated domain ontology as the primary ontology and the domain expert ontology as the secondary ontology, for the comparison process. The results of the enhanced similarity measure are validated against a standard dataset.

5.1 CROSS ONTOLOGY MEASURE

A method used for the integration of one or more ontologies to form a more generic ontology is considered as the cross ontology. A major component of creating a cross ontology is the semantic similarity measure. The proposed system focuses only on deriving a cross ontology measure. The two ontologies used here are the automated domain ontology and domain expert ontology.

Automated domain ontology: The automated domain ontology is associated to the domain, from which the ontology content is constructed. The ontology is constructed based on the user’s interest level and automated relationship between the concepts. The main part of the automated domain ontology is the automatically generated concept maps. The concept map helps to retrieve the relevant topic for the users from the e-learning system according to the user’s need and relevance. The concept map is a networked structure of concepts, which are interrelated to each other through an association.

Domain expert ontology: The domain expert ontology is defined over the set of topics by analysing the topics manually and creating the ontology. The main advantage of the ontology is that the construction is assisted by domain experts. So, it
guarantees the precision and efficiency of the ontology to the maximum level. Though the advantages are high for domain expert’s ontology, it has not been used frequently as the time for constructing domain ontology is very high and there are limitations in processing the deep rooted nodes. The proposed approach processes the automated domain ontology with “database” domain and the same domain is used by the domain experts to construct the manual ontology.

5.2 AN APPROACH TO FIND CROSS SIMILARITY MEASURE

The proposed approach leads to an algorithm to calculate the cross similarity between the two ontologies. The cross similarity measure is a measure to explore the contribution of a particular concept for a specific user query. A procedural approach is used for the cross similarity calculation. The process concentrates on the concepts in both the ontologies. From recent investigations by researchers (Mohammed Nazim Uddin et al, 2013) it is stated that, the concept of similarity calculation can improve the efficiency of retrieving the documents from the database. The cross similarity calculation goes through three main phases:

- Concept extraction
- Feature extraction and
- Cross similarity calculation

![Figure 5.1: Steps in finding cross similarity measure](image)

Figure 5.1: Steps in finding cross similarity measure
According to the three phases, a concept is processed and the relevant documents are extracted from the database using the cross similarity measures. Figure 5.1 shows the different processes in finding the cross similarity measure for a particular concept, which is triggered from the two ontologies based on the user’s input query. Now, let us discuss the process in detail.

5.2.1 CONCEPT EXTRACTION BASED ON USER QUERY

The initial phase in the proposed cross similarity measure calculation is the extraction of concepts from the ontologies. The concept extraction starts by accepting the query, which is given by the user as input. The query is a keyword, which is given by the user to extract the relevant information for the user from the e-learning system. Initially, the user query is processed using the automated domain ontology, which acts as the primary ontology for the proposed approach. The user query is compared with each concept in the automated domain ontology and a concept is selected from the ontology, which is more similar to the query keyword. Similarly, the query keyword is processed with each concept in the domain expert ontology. So, the initial phase generates a two element set for the next phase.

Step1. Select automated domain ontology
Step2. Select Domain expert ontology
Step3. Accept user query q
Step4. Compare q with each element in automated domain ontology
Step5. Select concept $c_a$
   where $c_a \cong q$
Step6. Compare q with each element in domain expert ontology
Step7. Select concept $c_d$
   where $c_d \cong q$
Step8. Construct concept set,
   \[ C_{set} = \{ c_a, c_d \} \]
Step9. End

Figure 5.2: Concept extraction algorithm
In our proposed approach, we consider the automated domain ontology as the primary ontology and the domain expert ontology as the secondary ontology. The extracted concepts are then stored in a set represented as concept set, $C_{set}$. The next phase of the cross similarity measure starts by accepting the $C_{set}$ as input. In the above algorithm, $q$ is the input query, $c_a$ is the concept in the automated domain ontology and $c_d$ is the concept in the domain expert ontology.

5.2.2 FEATURE EXTRACTION FROM CONCEPT SET

The next major phase of the proposed approach is the extraction of features from the concepts. The concept set is selected and the concepts are used for extracting the features. The features of the concepts are used for the calculation of the similarity measures. The similarity measure uses the comparison values of the features also to get more precise output. The higher the comparison value of the features the more is the weightage of relationships between the query and concepts. The neighborhood of each concept is considered as the features. The neighborhood is classified into two groups, called the representatives (super nodes) and properties (sub nodes). The proposed approach defines the features of concepts as two different sets as shown in Figure 5.3, one set containing the representatives of the concept and the other set containing the properties of the concepts.

![Figure 5.3: Representatives and properties](image)
5.2.2.1 REPRESENTATIVES OF CONCEPT SETS

The concept of representatives originated from the fact that a concept under consideration is a representative of its super classes. So, identification of the super classes of the concept can improve the results of the users query search, i.e. by incorporating the root nodes of the concept, which is extracted, can make the inter relationship stronger. Initially, the concept set, $C_{set}$ is selected for the processing. Then, the concepts from the concept set are selected. A bottom up tree traversal approach is applied to both the ontologies by setting the concept as the pivot point. All the super nodes of the selected concepts are extracted and are called as hypernyms. The output of the tree traversal process is a set of concepts,

$$c_{tree}^{\text{bottom-up}} = \{c_1, c_2, c_3, \ldots, c_n\}$$

The concepts extracted from both the ontologies are analysed and the distinct concepts are selected as the representatives of the concepts. Now the concept set containing representatives can be defined as

$$C_{representatives} = \{(c_a, c_d) / c_a \in O_{automated}, c_d \in O_{domain}\}$$

5.2.2.2 PROPERTIES OF CONCEPT SET

Another set contains the node which is a measure of part of relationship between the concepts in the ontology. These nodes are called the properties of the concept (sub nodes). A top down tree traversal approach is applied to both the ontologies by setting the concept as the pivot point. All sub nodes of the selected concept are extracted and are called as hyponyms. The output of the traversal process is a set of concepts

$$c_{tree}^{\text{top-down}} = \{c_1, c_2, c_3, \ldots, c_n\}$$

The concepts extracted from both the ontologies are analysed and the distinct concepts are selected as the properties of the concepts. Now the concept set containing properties can be defined as

$$C_{properties} = \{(c_a, c_d) / c_a \in O_{automated}, c_d \in O_{domain}\}$$
Once all the traversal processes are completed, a new set is formed with the super nodes and sub nodes of the concepts. These concepts are then used for the similarity measure calculation.

Step1. Select *Automated domain ontology*

Step2. Select *Domain expert’s ontology*

Step3. Select a concept from \( C_{\text{set}} \)

Step4. For the selected concept apply bottom up traversal and extract super nodes

Step5. Set \( C_{\text{bottom-up}}^{\text{tree}} = \{c_1, c_2, \ldots, c_n\} \)

Step 6. Construct the distinct representative set

\[ C_{\text{representatives}} = \{(c_a, c_d) / c_a \in O_{\text{automated}}, c_d \in O_{\text{domain}}\} \]

Step7. For the selected concept apply top down traversal and extract sub nodes

Step8. Set \( C_{\text{top-down}}^{\text{tree}} = \{c_1, c_2, \ldots, c_n\} \)

Step9. Construct distinct properties set

\[ C_{\text{properties}} = \{(c_a, c_d) / c_a \in O_{\text{automated}}, c_d \in O_{\text{domain}}\} \]

Step10. End

Figure 5.4: Feature extraction algorithm

### 5.2.3 CROSS SIMILARITY MEASURE CALCULATION

The cross similarity measure (ConSim) is the similarity calculation of the representatives and properties between the automated domain ontology and the domain expert ontology. It is equally important to calculate the similarity of the representatives and properties to improve the accuracy for considering a concept to be relevant for a given input query. So, according to the proposed approach, the features (representatives and properties) of the query concept are extracted from the automated and domain ontology. We initially calculate the similarity of the representatives.

Let \( C_{\text{representatives}} \) be the set of representatives extracted for a particular concept in the concept set \( C_{\text{set}} \).
Let \( C_{represents} = \{(c_a, c_d) \mid c_a \in O_{automated}, c_d \in O_{domain}\} \)

Let \( rc_a \) represents the set of representatives from \( O_{automated} \) and \( rc_d \) represents the set of representatives from \( O_{domain} \).

\[
rc_a = \frac{c_a \cap c_d}{c_a}
\]

\[
rc_d = \frac{c_a \cap c_d}{c_d}
\]

Thus, each element \( c_a \) in \( O_{automated} \) and each element \( c_d \) in \( O_{domain} \), which are selected as representatives are compared to each other by the function defined as,

\[
R_{similarity} = \sqrt{\frac{rc_a^2 + rc_d^2}{2}}
\]

The values in the above expression define the similarity between the representatives from \( O_{automated} \) and representatives from \( O_{domain} \) separately. According to the \( R_{similarity} \) equation, similarity values of each and every representative of the concept \( c \) are calculated.

Now, we move on to the calculation of the similarity between properties defined for the concept \( c \) over the two ontologies. Let \( C_{properties} \) be the set of representatives extracted for a particular concept in the concept set \( C_{set} \).

\[
C_{properties} = \{(c_a, c_d) \mid c_a \in O_{automated}, c_d \in O_{domain}\}
\]

Let \( pc_a \) represents the set of properties from \( O_{automated} \) and \( pc_d \) represents the set of properties from \( O_{domain} \).

\[
pc_a = \frac{c_a \cap c_d}{c_a}
\]

\[
pc_d = \frac{c_a \cap c_d}{c_d}
\]

Each element \( c_a \) is from \( O_{automated} \) and each element \( c_d \) is from \( O_{domain} \), which are selected as properties and compared with each other by the function defined as,
The property similarity calculation is also calculated similar to the representative calculation. All the similarity values are selected and subjected for the concept similarity calculation. The concept similarity value is defined as,

\[ C_{sim} = \sqrt{\frac{\alpha R_{similarity}^2 + \beta P_{similarity}^2}{2}} \]

where, \( \alpha = \frac{c_a}{c_a + c_d}, c_a, c_d \in C_{representatives} \) ; \( \beta = \frac{c_a}{c_a + c_d}, c_a, c_d \in C_{properties} \)

According to the above expression, \( C_{sim} \) values for all concepts in the primary domain are calculated and one concept with a high \( C_{sim} \) value and its associated concepts are selected as the most relevant results for the input query. The associated concepts are the features of the concepts which have a similar \( C_{sim} \) value with that of the selected concept.

The cross similarity measure is the value that specifies the importance of a concept, which is retrieved as a result of input query. The cross similarity measure plays an important role in the retrieval of documents from an e-learning system.

Step 1. Select \( C_{representatives} \)

\( C_{representatives} = \{ (c_a, c_d) / c_a \in O_{automated}, c_d \in O_{domain} \} \)

Step 2. For each element in \( C_{representatives} \) compute representative similarity

\[ R_{similarity} = \sqrt{\frac{r_{c_a}^2 + r_{c_d}^2}{2}} \]

where, \( r_{c_a} = \frac{c_{a} \cap c_{d}}{c_{a}} ; r_{c_d} = \frac{c_{a} \cap c_{d}}{c_{d}} \)

Step 3. Select \( C_{properties} \)

\( C_{properties} = \{ (c_a, c_d) / c_a \in O_{automated}, c_d \in O_{domain} \} \)

Step 4. For each element in \( C_{properties} \) compute property similarity

\[ P_{similarity} = \sqrt{\frac{p_{c_a}^2 + p_{c_d}^2}{2}} \]
where, $p_{ca} = \frac{c_a \cap c_d}{c_a}$; $p_{cd} = \frac{c_a \cap c_d}{c_d}$

Step 5. Define concept similarity $C_{sim}$

$$C_{sim} = \sqrt{\frac{\alpha^2 \text{similarity} + \beta^2 \text{similarity}}{2}}$$

where, $\alpha = \frac{c_a}{c_a + c_d}, c_a, c_d \in C_{representatives}$

$$\beta = \frac{c_a}{c_a + c_d}, c_a, c_d \in C_{properties}$$

Step 6. Calculate $C_{sim}$ for all the features of the concept in $O_{automated}$ and $O_{domain}$.

Step 7. Select the concept $c$ with maximum $C_{sim}$ value.

Step 8. Select all associated concepts closer to the $C_{sim}$ measure.

Step 9. Retrieve documents based on selected concepts.

Step 10. End

Figure 5.5: Algorithm cross similarity measure

5.3 EXPERIMENTAL EVALUATION OF THE PROPOSED MEASURE

The derived similarity measure is evaluated against a standard dataset provided by the Intelligence laboratory, Technical University of Crete and its superiority over the existing hybrid similarity measures is shown. This similarity measure is then applied for the comparison of the two ontologies discussed in section 5.3.2 and the evaluation results are also discussed.

5.3.1 COMPARATIVE EVALUATION OF THE DESIGNED CROSS ONTOLOGY MEASURE

The superiority of the proposed method can be determined by comparing it with some other standard datasets. So, a comparison analysis of the proposed approach on a medical database is performed. The ontologies, which are taken into account, are the WordNet and MeSH. WordNet (Hirst G. and St.Onge, 1998) is a lexical database for the English language. It groups English words into sets of
synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets (http://wordnet.princeton.edu/). MeSH (Medical Subject Headings) is a taxonomic hierarchy of medical and biological terms suggested by the U.S. National Library of Medicine (NLM) (http://en.wikipedia.org/wiki/Medical_Subject_Headings). Let us consider the samples of the MeSH ontology and WordNet ontology over the OWL format.

Dementia–Atopic Dermatitis

WordNet–Dementia

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<Term>Dementia
<Hypernyms>Insanity,, mental deterioration of organic or functional origin</Hypernyms>
<Hypnyms>Alcoholic dementia, alcoholic amnestic disorder, Korsakoff's psychosis, Korsakoff's syndrome, Korsakoff's psychosis, Korsakoff's syndrome, polymyelitic psychosis,, dementia, dementedness</Hypnyms>
</Term>

MeSH–Atopic Dermatitis

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<Term>Atopic dermatitis
<Hypernymns>A dermatitis that is a chronically relapsing inflammatory allergic response located in the skin that causes itching and flaking.. A dermatitis that is a chronically relapsing inflammatory allergic response located in the skin that causes itching and flaking.</Hypernymns>
<Hypnymns>Benign &apos;apos;s prurigo, allergic dermatitis, Atopic dermatitis NOS (disorder), allergic dermatitis, Atopic neurodermatitis (disorder), Allergic (intrinsic) eczema (disorder), Atopic dermatitis (disorder), Benign &apos;apos;s prurigo (disorder), Atopic dermatitis</Hypnymns>
</Term>

Figure 5.6: Sample: WordNet and MeSH

Figure 5.6 shows the sample XML file generated for WordNet and MeSH ontology through provision of two queries. These XML files are generated by providing the input as, “dementia” for WordNet and “Atopic Dermatitis” for MeSH. Similarly, a number of queries are given simultaneously to WordNet and MeSH and the different similarity measures are applied to these extracted keywords (based on the query). The standard cross ontology similarity algorithms, which are used to compare with the proposed cross similarity method, are X-similarity measure (2006) and Rodriguez M. A. et al. measure (2003) which are discussed in section 2.4.2.2. The results obtained from these algorithms are given in Table 5.1.
Table 5.1: Comparative Analysis of similarity measures with human score

<table>
<thead>
<tr>
<th>Query Keyword</th>
<th>Similarity Measures</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WordNet</td>
<td>MeSH</td>
<td>Human Score</td>
<td>X-Similarity measure</td>
</tr>
<tr>
<td>Anemia</td>
<td>Appendixitis</td>
<td>Tricuspid Atresia</td>
<td>0.0313</td>
<td>0</td>
</tr>
<tr>
<td>Meningitis</td>
<td>Mental Retardation</td>
<td>0.0313</td>
<td>0.025</td>
<td>0.0083</td>
</tr>
<tr>
<td>Sinusitis</td>
<td>Atopic Dermatitis</td>
<td>0.0625</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dementia</td>
<td>Bacterial Pneumonia</td>
<td>0.1563</td>
<td>0.113</td>
<td>0</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Patent Ductus</td>
<td>0.1563</td>
<td>0.122</td>
<td>0</td>
</tr>
<tr>
<td>Congenital Heart</td>
<td>Acquired Immunodeficiency</td>
<td>0.1563</td>
<td>0.084</td>
<td>0</td>
</tr>
<tr>
<td>Otitis Media</td>
<td>Infantile Colic</td>
<td>0.1563</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>Hyperkalemia</td>
<td>0.1563</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sepsis</td>
<td>Neonatal Jaundice</td>
<td>0.1875</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>Asthma</td>
<td>0.375</td>
<td>0.07</td>
<td>0.0119</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>Hyperthyroidism</td>
<td>0.4063</td>
<td>0.387</td>
<td>0</td>
</tr>
<tr>
<td>Sarcoidosis</td>
<td>Tuberculosis</td>
<td>0.4063</td>
<td>0.387</td>
<td>0</td>
</tr>
<tr>
<td>Iron Deficiency Anemia</td>
<td>Sickle Cell Anemia</td>
<td>0.4375</td>
<td>0.14</td>
<td>0.0117</td>
</tr>
<tr>
<td>Adenovirus</td>
<td>Rotavirus</td>
<td>0.4375</td>
<td>0.16</td>
<td>0.0187</td>
</tr>
<tr>
<td>Lactose Intolerance</td>
<td>Irritable Bowel Syndrome</td>
<td>0.4688</td>
<td>0.047</td>
<td>0.0057</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetic Nephropathy</td>
<td>0.5</td>
<td>0.205</td>
<td>0.0183</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>Hepatitis C</td>
<td>0.5625</td>
<td>0.42</td>
<td>0.016</td>
</tr>
<tr>
<td>Psychology</td>
<td>Cognitive Science</td>
<td>0.5938</td>
<td>0.25</td>
<td>0.0083</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>Failure to Thrive</td>
<td>0.625</td>
<td>0.043</td>
<td>0.0143</td>
</tr>
<tr>
<td>Urinary Tract Infection</td>
<td>Pyelonephritis</td>
<td>0.6563</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Myocardial</td>
<td>Myocardial Ischemia</td>
<td>0.7188</td>
<td>0.47</td>
<td>0</td>
</tr>
<tr>
<td>Headache</td>
<td>Migraine</td>
<td>0.75</td>
<td>0.042</td>
<td>0</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>Neoplasm</td>
<td>0.75</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td>Breast Feeding</td>
<td>Lactation</td>
<td>0.8438</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pain</td>
<td>Ache</td>
<td>0.875</td>
<td>1</td>
<td>0.0217</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>Nutritional Deficiency</td>
<td>0.875</td>
<td>1</td>
<td>0.1433</td>
</tr>
<tr>
<td>Down Syndrome</td>
<td>Trisomy 21</td>
<td>0.875</td>
<td>1</td>
<td>0.1467</td>
</tr>
<tr>
<td>Antibiotic</td>
<td>Antibacterial Agents</td>
<td>0.9375</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>Varicella</td>
<td>Chickenpox</td>
<td>0.9688</td>
<td>1</td>
<td>0.2477</td>
</tr>
<tr>
<td><strong>Correlation Values</strong></td>
<td></td>
<td><strong>0.9923</strong></td>
<td><strong>0.6974</strong></td>
<td><strong>0.5701</strong></td>
</tr>
</tbody>
</table>
The analysis from Table 5.1 shows that the similarity values obtained using the proposed approach for almost all the queries are closer to the human score, while other algorithms have failed to provide similarity values for all the queries. The results from the different algorithms for different pairs of queries are shown in Figures 5.7, 5.8 and 5.9.

Figure 5.7: Comparison between Human score and Rodriguez et al. measure

Figure 5.8: Comparison between Human score and X-Similarity measure
This comparative analysis in Figure 5.10 shows that the proposed approach gives importance to each and every sub data associated with the given query. Correlation coefficient is a standard method used to compare the similarity between two sets of
data. The similarity values obtained using each method is compared with the human score using correlation coefficient. The average similarity value obtained using the proposed similarity measure is 0.898, while the method proposed by Rodriguez M.A. et al. has given 0.57 and X- similarity measure was 0.697. Thus, it can be stated that, the proposed approach (ConSim) is more effective in calculating the cross similarity measure irrespective of any database.

5.3.2 COMPARATIVE EVALUATION OF THE DESIGNED CROSS ONTOLOGY MEASURE FOR TWO ONTOLOGIES

The proposed cross ontology measure is developed based on two main ontologies: the automated domain ontology and the domain expert ontology. The data from the automatically generated concept maps and manually generated concepts are given to the Protege tool for depicting the ontology. An important factor in selecting Protege as the tool is because it generates a clear and evident OWL for the proposed ontology. The concepts and their association values are embedded on the OWL language with the help of Protege tool. The super nodes and the sub nodes are also clearly depicted using the Protege tool. The automated domain ontology is considered as the primary ontology and the domain expert ontology is considered as the secondary ontology for the proposed cross ontology similarity measure based method.

5.3.2.1 CROSS ONTOLOGY MEASURE FOR DIFFERENT QUERY WORDS

The following section describes an analysis of the performance of the cross ontology similarity measure. The automated ontologies are generated using association rule mining and prefix span method. The domain expert ontology is generated manually. The comparisons are divided into different cases and discussed in what follows:
Case 1: Comparison of the existing automated ontology with the domain expert ontology.

The evaluations are done based on two queries (keywords), i.e. two queries will be selected from the two ontologies (automated ontology and domain expert ontology) and their similarity is measured on the basis of cross ontology similarity measure. Figure 5.8 shows a sample similarity measure evaluated for a single pair of keywords. Database from automated ontology and database_design from domain ontology are given as input and the cross similarity measure is evaluated and found to be 1.732. Table 5.2 shows five pairs of query keywords and their similarity values. The optimum similarity measure obtained is 1.862 for the proposed approach.

Figure 5.11: Cross similarity evaluation for a single pair.
Table 5.2: Cross similarity measure for ontologies generated using existing approach

<table>
<thead>
<tr>
<th>Automated ontology (Existing approach)</th>
<th>Domain expert ontology</th>
<th>Similarity measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorting_data</td>
<td>Sorted_Files</td>
<td>1.861898673</td>
</tr>
<tr>
<td>condition_partition</td>
<td>disk_space_management</td>
<td>1.861898673</td>
</tr>
<tr>
<td>Database</td>
<td>Database_design</td>
<td>1.732050808</td>
</tr>
<tr>
<td>data_disks</td>
<td>storing_data</td>
<td>1.558578412</td>
</tr>
<tr>
<td>relational_algorithm</td>
<td>relational_model</td>
<td>1.861898673</td>
</tr>
</tbody>
</table>

Case 2: Comparison of the automated domain ontology constructed using association rule mining with the domain expert ontology.

Here the evaluation is done based for case 2 and the measures are shown in Table 5.3 which represents the similarity measure obtained for the proposed approach by providing different set of query keywords. The optimum similarity measure obtained is 2.141 for the proposed approach.

Table 5.3: Cross similarity measure for ontologies generated using association rule mining

<table>
<thead>
<tr>
<th>Automated ontology (Association rule mining)</th>
<th>Domain expert ontology</th>
<th>Similarity Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>relational_database</td>
<td>relational_value</td>
<td>1.943650632</td>
</tr>
<tr>
<td>Database</td>
<td>database_design</td>
<td>2.141603643</td>
</tr>
<tr>
<td>transaction_management</td>
<td>Sorting</td>
<td>1.943650632</td>
</tr>
<tr>
<td>query_data</td>
<td>query_evaluation</td>
<td>1.219631092</td>
</tr>
<tr>
<td>relational_model</td>
<td>relational_database</td>
<td>2.141603643</td>
</tr>
</tbody>
</table>

Case 3: Comparison of the automated domain ontology constructed using PrefixSpan algorithm with the domain expert ontology.

Here the evaluation is done based for case 3. Table 5.4 represents the similarity measure obtained for the PrefixSpan based approach by providing different set of query keywords. The optimum similarity measure obtained is 2.2125 for the proposed approach.
Table 5.4: Cross similarity measure for ontologies generated using PrefixSpan mining

<table>
<thead>
<tr>
<th>Automated ontology (PrefixSpan Mining)</th>
<th>Domain expert ontology</th>
<th>Similarity Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>relational_operators</td>
<td>relational_algebra</td>
<td>2.212565825</td>
</tr>
<tr>
<td>relational_model</td>
<td>relational_database</td>
<td>2.212565825</td>
</tr>
<tr>
<td>data_disks</td>
<td>disk_space_management</td>
<td>1.558578412</td>
</tr>
<tr>
<td>structured_indexing</td>
<td>Redundancy</td>
<td>1.219631092</td>
</tr>
<tr>
<td>data_base</td>
<td>storing_data</td>
<td>2.15245797</td>
</tr>
</tbody>
</table>

Now, a cross comparison is performed by incorporating all the three approaches and by selecting similar queries as input. This comparison will help to identify an efficient approach based on similarity.

Table 5.5: Comparison of cross similarity measures

<table>
<thead>
<tr>
<th>Automated Ontology (Query keyword1)</th>
<th>Domain Ontology (Query keyword2)</th>
<th>Cross similarity measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Existing method Case 1</td>
</tr>
<tr>
<td>Data_base</td>
<td>Storing_data</td>
<td>1.219631092</td>
</tr>
<tr>
<td>Database</td>
<td>Database_design</td>
<td>1.732050808</td>
</tr>
<tr>
<td>Relation_model</td>
<td>Relational_database</td>
<td>1.861898673</td>
</tr>
<tr>
<td>relational_operators</td>
<td>relational_algebra</td>
<td>1.732050808</td>
</tr>
</tbody>
</table>

Table 5.5 represents cross similarity measures obtained for three different approaches. The analysis from the Figure 5.11 shows that, the cross similarity measure based on the proposed approaches has higher similarity measure values, which eventually shows that, the proposed method has higher efficiency over the existing approach.
Figure 5.12: A comparison of different cross ontologies measures

Figure 5.12 represents the comparison of similarity values obtained by comparing the manual ontology with the ontologies obtained using prefix span method, association rule mining and an existing BMI method. The results are obtained by providing four pairs of similar keywords to each cross ontology. The analysis from the graph shows that the proposed approach based on prefix span method outperforms the other two methods.

5.4 CONCLUSIONS

The proposed approach uses multiple ontologies for the evolution of the concept for a particular domain. The use of more than one ontology improves the performance and accuracy of the proposed method. The experimental result shows that the system used in this work provides feasible results and outperforms the existing methods. The existing hybrid similarity measures have been discussed in Chapter 2 and compared with the new enhanced concept similarity measure (ConSim). The derived concept similarity measure is compared with the terms of standard ontologies like WordNet and MeSH. Experimental results proved that the derived measure is closer to the Human Scores published by Intelligent System
laboratory, TUC, when compared with the other hybrid measures. The correlation coefficient of the ConSim measure was 0.89 when compared to the measures of Rodriguez and X-similarity which are 0.55 and 0.67 respectively. The results indicate that the ConSim similarity measure outperforms the other hybrid measures. This measure has been used to compare the automated ontologies with the domain expert ontology. The similarity measures proved that the prefix scan method of automating ontology is better than the association rule mining method, which in turn is better than the existing method.