3. Organization of Concepts and Content in Ontology

The knowledge about academic subjects, related to topics or concepts may be organized and described as relations in Concept Map. SKOS (Simple Knowledge Organization System) allows building the taxonomical relations among various concepts under a domain. Use of such tools benefits applications such as e-Learning where educational concepts and taxonomical relations among their related concepts need to be described. They are simpler to build as compared to other expensive ontologies and more expressive to describe relationships among contextual concepts. However both have their limitations when used for gathering resources from the Web and organizing them for later retrieval.

This chapter proposes the design of Concept Ontology taking advantages of structures from Concept Map and SKOS. The proposed Concept Ontology consists of two ontologies, Domain Ontology (DO) and Content Fragment Ontology (CFO). The schema of the Concept Ontology has been designed to represent DO, CFO and the relationships among their classes. The DO contains the concepts and domains linked through various relations they possess, whereas the CFO consists of relevant resources (called fragments) linked to their corresponding relevant concepts in the DO. The knowledge inferred from these ontologies can be utilized by different Information Retrieval sub-processes. Term Expansion is one such sub-process of IR process which can get benefit from the Concept Ontology.

3.1 Introduction

Nowadays, much of the research in the area of e-Learning has focused on the organization of learning content so that the learning resources can be reused which may lead to savings in time and money. This consequently enhances the quality of digital learning experiences (Duval & Hodgins, 2003). For such a learning experience, the organization mechanism must enable the sharing, reuse and identification of the content (resources) in a meaningful and useful way. To achieve such organization, important decomposition techniques found in AI methodology for building software systems can be applied (Gruber T. R., 1991) here also in building the Concept Ontology. According to these methodologies, it is required to: i) separate knowledge
from programs with a declarative knowledge representation language, ii) identify general classes and relations underlying application specific facts, and organize knowledge to enable inheritance from these constructs, and iii) characterize general problem solving tasks (classification) and classes of inference (subsumption), and design corresponding methods and algorithms.

Ontology provides a solution towards the issue regarding the organization of concepts and content. It defines (specifies) concepts, relationships, and other distinctions that are self-explanatory for modeling a domain (Gruber T. R., 1993), (Gruber T. R., 1995). Ontology provides an efficient solution for defining semantic relations among various terms or concepts. However, the relations among concepts belonging to academic subjects are abstract and therefore need a different approach for building the knowledge base. Instead of using hypernyms, hyponyms, synonyms or antonyms from the WordNet which forms a taxonomic hierarchy of natural language terms, sub-processes of Information Retrieval such as content gathering/retrieval from Web, query term expansion and so on use concepts consisting of abstract relations. This is because the technical terms usually do not have synonyms, hyponyms etc. rather they have sub concepts, super concepts, sibling concepts and other associated relations that define them.

The requirement of such concept based ontology can further be explained with the following example. A resource or a web document usually does not consist of only a single concept; rather it consists of many related concepts from the same or other domains. A single concept cannot be explained alone either, it requires other related concepts that subsequently assist in knowing and understanding the concept. This argument implies that a single resource can be used to explain multiple concepts. Therefore, while building a knowledge base, the system must possess enough intelligence to identify different concepts in a resource and also that, how relevantly the resource defines each of the contained concepts. This information if gets identified by the system then it is required to be organized in the knowledge.

A design of ontology with such an organization has been proposed in this chapter. It incorporates above requirements pertaining to semantic linking of the abstract concepts and their linking to the corresponding relevant resources as well.
An important usage of the Concept Ontology employed in the thesis is Term Expansion. The expansion of a concept using the Concept Ontology can prove beneficial to the retrieval and ranking processes of IR system, as explained later in the chapter.

Query Expansion in general and Term Expansion in particular is one of the important tasks of IR apart from crawling and searching (Wu, Ilyas, & Weddel, 2011). Query Expansion (QE) is regarded as an IR process which consists of selecting and adding terms to the user's query with the goal of retrieving more relevant documents or resources and thus improving retrieval performance (Vechtomova, 2009). Here, the thesis specifically uses ‘Term Expansion’ instead of QE as this process is used not only at the time of searching but during the crawling and ranking processes as well. Moreover, unlike QE which usually deals with multiple terms, Term Expansion works on single concept or topic; nevertheless, a topic or concept may consists of compound words.

### 3.2 Related Work

The educational ontologies are majorly designed for Learning Objects (LO) as is evident from the literature. Researchers (Wang, Fang, & Fan, 2008), (Kiu & Lee, 2005), (Wang S., 2008) have proposed ontology designs for LOs so as to make the objects available for reusing and sharing purpose. These LOs however, are required to be designed according to the application need, e.g., a course specific pedagogical based content sequencing. A research (Jovanović, Gašević, & Devedžić, 2006) rather redefined LOs by further decomposing them at finer granularity units called fragments.

A domain specific ontology presented in the research (Sosnovsky & Gavrilova, 2005) showed a design for organizing concepts only. The framework presented by (Ghaleb, Daoud, Hasna, Jaam, & El-Sofany, 2006) is an application-specific model designed to represent a semantic relation within an e-Learning system. A similar work by Fernandez et al. (2011) investigated the process of selection, extraction and reorganization of the content from Semantic Web information sources rather than the WWW web resources.
Concept Ontology as proposed in this chapter relates various concepts lying under a domain with structured in a taxonomical hierarchy under a domain, quite similar to SKOS (Isaac & Summers, 2009). Besides it also describes relationships among concepts as it is described by Concept Maps (Novak J. D., 2010). Concept maps are graphical tools for organizing and representing knowledge (Novak & Cañas, 2006-01 Rev 01-2008). Sometimes they are also called semantic units as they describe the meanings of relations among concepts and in a way describe the concept too. Concept maps are developed on the strong theoretical foundation which provides many benefits in the field of education (Novak J. D., 2010). The work on constructing ontology to concept maps (Graudina & Grundspenkis, 2008) and vice versa (Simón, Luigi, & Alejandro, 2007) has been implemented in past, to serve different needs. The research in past (Kekalainen & Jarvelin, 1998) has shown best precision scores when a query term is expanded with its narrower concepts, associative concepts and synonyms.

One such initiative to structure the concepts and the available resources is Topic Map, an ISO/IEC 13250. It consists of a collection of topics, each of which represents some concept. Topics are related to each other by associations, which are typed n-ary combinations of topics. A topic may also be related to any number of resources by its occurrences (Ahmed & Moore, 2005). However, the Topic Maps do not come with a pre-defined ontology.

According to Carpineto & Romano (2012) QE techniques can be classified into five main groups according to the conceptual paradigm used for finding the expansion features. These five groups are linguistic analysis, corpus-specific statistical approaches, query-specific statistical approaches, search log analysis and Web data. Web data (Arguello, Elsa, Callan, & Carbonell, 2008) consisting of structured portals such as anchor text, Wikipedia, Social Network Sites etc. can effectively be used to extract contextual and related keywords to the original query. Search log analysis is one of the widely researched methods to find relevant terms associated to a query (Kekalainen & Jarvelin, 1998).

Search logs contain user queries, followed by the URLs of Web pages that are clicked by web users. A number of techniques such as mining user logs to extract
information from users’ interaction (Cui, Wen, Nie, & Ma, 2003) and clickthrough to determine association between query and documents (Billerbeck, Scholer, Williams, & Zobel, 2003). The technique may encode implicit relevance feedback (Rocchio, 1971) which statistically assess the documents and use them to adjust the weights of terms in the original query and/or to add words to the query (Amati, Carpineto, & Romano, 2001). Besides the availability of large-scale search logs, the problems associated with search logs are noise, incompleteness, sparseness, and the volatility of web pages and query (Xue, et al., 2004).

Query specific-techniques (Bai, Nie, Bouchard, & Cao, 2007) are based on preprocessing the top retrieved documents for filtering out irrelevant features prior to the utilization of a term-ranking function (Billerbeck & Zobel, 2005). Since the documents are retrieved through analyzing the associated queries with them, the method is able to retrieve and suggest more relevant documents. However, the associated issues of cold start problem and rare queries do exist (Broder, et al., 2009).

Another classic approach to QE is document corpus statistical analysis (Gauch, Wang, & Rachakonda, 1999). It analysis correlations between pairs of terms by exploiting term co-occurrence, either at the document level, or at the restricted contexts such as paragraphs, sentences, or small neighborhoods. Concept terms (Qiu & Frei, 1993) and term clustering (Bast, Majumdar, & Weber, 2007) used in the approach are two classical strategies.

Linguistic analysis approach used for expanding an original query uses the knowledge structure usually represented as thesaurus, taxonomy or ontology. Thesaurus such as WordNet can be used to disambiguate (Navigli, 2009) the context or sense of the given query (Hidalgo, Rodríguez, & P´erez, 2005), whereas the query with a specified domain can be expanded more reliably using the taxonomy or ontology (Lee, Tsai, & Wang, 2008). This approach alleviates the need of users’ feedback and click logs as these may give biased results (Grootjen & Weide, 2006); (Bhogal, Macfarlane, & Smith, 2007).
3.3 The Proposed Concept Ontology Framework

A design of a domain-specific Concept Ontology has been proposed in this section. The ontology has been designed in such way that the links among the concepts do not only describe their generalized and specialized relationships but also links to the content which describes them. This organization could help in effectively expanding any concept for better contextual search results. The proposed Concept Ontology framework is illustrated in Figure 3-1. The ontology framework encompasses the two level ontology designs linked through relevance relations. The two levels are Domain Ontology and Content Fragment Ontology which are described following in detail. The schema consisting of class definitions and relations among them is explained in Concept Ontology. Following are the list of URIs and their corresponding prefixes used in the subsequent sections to describe the different ontologies.

![Concept Ontology Framework](image)

**Figure 3-1: Concept Ontology Framework**

```xml
<!DOCTYPE rdf:RDF [  
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >  
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >  
  <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >  
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >  
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >  
  <!ENTITY core "http://www.cs.du.ac.in/2011/10/coewr/core.owl#" >  
]>
```
3.3.1 Domain Ontology

The knowledge base in Domain Ontology consists of various concepts under a domain. A single concept may belong to multiple domains and at the same time multiple concepts from different domains may also be linked. This arrangement is essential to integrate multiple domains in order to define better semantic relations, and is taken care in the Concept Ontology design. Since the Domain Ontology represents

![Diagram of concept ontology]

Figure 3-2: A segment of concept ontology illustrating relations among concepts under a domain
most of the relations among the concepts and is integral part of the system, it is required to be constructed carefully by using expert’s knowledge following multiple iterations. The Domain Ontology has been developed manually using protégé editor. It consists of concepts and relations under different domains.

The domain ontologies form the basis for determining semantic relation among various concepts and thus consequently used to compute the empirical semantic distance between the pair of concepts. No doubt, they need to be carefully designed to represent a correct and factual relationship among different concepts. Although they are expensive to build, constructing ontology is a one-time process. Domain Ontology which depicts a structure of a broad topic or domain can be extended any time with ease, but it seldom requires any change. These ontologies can be reused later in number of applications. In the coming years the Semantic Web is expected to contain almost every topic of interest (Devedzic, 2004). At that time the topic expansion could be alternatively processed using knowledge from the GGG (Giant Global Graph) (Berners-Lee T., 2007).

Figure 3-2 shows a small segment of Concept Ontology consisting of few properties, a class and a few individuals in database domain. It shows that all individuals (DML, SQL, select and update) are of the type Concept which is a class. The properties hasSuperconcept and hasSubconcept have the domain and the range consisting of the class Concept (only the range is shown in the Figure 3-2). These properties (relations) resemble the SKOS design which uses broader and narrower properties to relate the concepts. Further, the properties, hasSuperconcept and hasSubconcept acts as inverse to each other. They also exhibit transitivity, which is shown at the bottom of the figure as the owl:TransitiveProperty property.

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#DML -->

<Concept rdf:about="#DML">
  <rdf:type rdf:resource="&owl;Thing"/>
  <hasSuperConcept rdf:resource="#SQL"/>
  <hasStatement rdf:resource="#delete"/>
  <hasSubConcept rdf:resource="#delete"/>
</Concept>
Figure 3-3: Database Domain Ontology
It is important to note here that the `rdf:resource` used in the concept definition above is different from the resource that has been used throughout the thesis which represents the web document or web resource as a resource for learning. The former ‘resource’ instead refers to an rdf resource, a part of ontology. More than 80 such concepts are defined in the Concept Ontology manually. They are defined using varied relations with other concepts.

Figure 3-3 illustrates the concepts in the ‘database’ domain ontology. Each concept in the ontology is defined with multiple relations. For an example, the concept DML and create in database ontology have been defined as following.

### 3.3.2 Content Fragment Ontology

The web resources are stored in fragment ontology as instances of the content fragments which are based on Content Fragment Classification (Verbert, Klerkx, Meire, Najjar, & Duval, 2004). The intent behind using this classification is to separate the structure from content which is an important step towards reusability, also suggested by Duval and Hodgins (2003) in their research agenda.

The content fragment is a most basic unit of content that represents independent and individual resources. Each fragment in CFO is linked to the concept(s) along with the relevance with respect to their corresponding concepts in Domain Ontology. CFO makes an early distinction between classes that describes discrete elements and continuous elements, the types of data formats. Subclasses of discrete elements are Graphics, Data files, Text, etc. whereas, Continuous elements have audio, video, animation etc. subclasses. Each of these subclasses constitutes various formats that are used in files/documents. The contents being separate from the context, allows an administrator an easy and simple maintenance tasks that involves updating ontology, addition of new classes etc. as per the need.

A unique and added benefit of using the CFO is that it stores the Social Semantic Relevance and Semantic Relevance (discussed in Chapter 4 and Chapter 5) and, the depth of the learning content in each web resource, associated with the concepts. Using this information, resource annotation is processed automatically during ontology augmentation (section 3.4.2) which is an advantage over the manually created metadata. This annotation helps to retrieve relevant resources from
Retrieving and Organizing Web Resources Semantically for Informal E-Mentoring

ontology which effectively reduces computational time. URLs of semantically relevant web resources to the domain concepts are organized according to their relation with their concepts in the ontology and hence named as Content Fragment Ontology.

Figure 3-4 illustrates the links in CFO classification. The owl representation of the above mentioned classes are as follows:

```xml
<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#Fragment -->

<owl:Class rdf:about="#Fragment"/>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#FragmentFormat -->

<owl:Class rdf:about="#FragmentFormat"/>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#ContinuousElements -->

<owl:Class rdf:about="#ContinuousElements">
    <rdfs:subClassOf rdf:resource="#FragmentFormat"/>
</owl:Class>
```

![Figure 3-4: Content Fragment Taxonomy](image)

Figure 3-4: Content Fragment Taxonomy
FragmentFormat is another class which defines the type of format each resource as a fragment possess.

3.3.3 Concept Ontology

The Concept Ontology defines the schema which encompasses Domain Ontology, Fragment Ontology and the relationships that exist between the entities of two ontologies. The relations amongst the concepts are expressed as different types of associations apart from the taxonomical relations such as super concepts and sub
concepts. The schema defines the Classes, Object properties and Data properties associated with the concepts.

3.3.3.1 A Concept in Concept Ontology

A Concept in the Concept Ontology is defined as a Class which describes an abstract concept under a domain. A Concept relates to other concepts as shown in Figure 3-3. At the same time it also defines the associated web resources that contain relevant content in varied formats to help in understanding the Concept. It also determines the depth of the content regarding the Concept and other related concepts through various relations (data properties and object properties) as shown in Figure 3-5 and Figure 3-6.

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#associatedTo -->

```xml
<owl:ObjectProperty rdf:about="#associatedTo">
  <rdfs:range rdf:resource="#Concept"/>
  <rdfs:domain rdf:resource="#Concept"/>
</owl:ObjectProperty>
```

The object property associatedTo is the only object property in Concept Ontology that has sub-properties. The sub-properties are hasArithmeticOperation, hasClause, hasFunction, hasOption, hasSetOperation and hasStatement. These properties are flexible in the

![Figure 3-5: Representation of ‘concept’ in Concept Ontology](image-url)
sense that many such properties can be added while constructing the Domain Ontology, depending on the abstract relationships between concepts.

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#belongsToDomain -->

<owl:ObjectProperty rdf:about="#belongsToDomain">
  <rdfs:domain rdf:resource="#Concept"/>
  <rdfs:range rdf:resource="#Domain"/>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasSibling -->

<owl:ObjectProperty rdf:about="#hasSibling">
  <rdfs:domain rdf:resource="#Concept"/>
  <rdfs:range rdf:resource="#Concept"/>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasSubConcept -->

<owl:ObjectProperty rdf:about="#hasSubConcept">
  <rdf:type rdf:resource="&owl;TransitiveProperty"/>
  <rdfs:range rdf:resource="#Concept"/>
  <rdfs:domain rdf:resource="#Concept"/>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasSuperConcept -->

<owl:ObjectProperty rdf:about="#hasSuperConcept">
  <rdf:type rdf:resource="&owl;TransitiveProperty"/>
  <rdfs:domain rdf:resource="#Concept"/>
  <rdfs:range rdf:resource="#Concept"/>
</owl:ObjectProperty>

Following is an example of the data property associated with the concept.

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasDesc -->

<owl:DatatypeProperty rdf:about="#hasDesc">
3.3.3.2 Linking Fragment in Concept Ontology

The web resources are crawled and gathered from the Web. The resources which are found relevant to the concepts are used as the fragments. These fragments are linked to the relevant concepts along with the information related to the relevance relation property. These links are used to augment the knowledge in the Concept Ontology programmatically as an automatic process.

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#inFragment -->

<owl:ObjectProperty rdf:about="#inFragment">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
  <rdfs:domain rdf:resource="#Fragment"/>
  <rdfs:range rdf:resource="#RelevanceRelation"/>
  <rdfs:domain>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#inFragment"/>
      <owl:someValuesFrom rdf:resource="#Fragment"/>
    </owl:Restriction>
  </rdfs:domain>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasURL -->

<owl:DatatypeProperty rdf:about="#hasURL">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
  <rdfs:domain rdf:resource="#Fragment"/>
  <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>
3.3.3.3 Adding Relevance Relation to the Concept

Relevance relation between a concept and a resource determines that how relevant is learning content in the resource related to the concept. This is a two array relationship and is handled as a special case in the ontology. Besides the empirical value of the relevance, it also determines the type of content depth of the resource (Basic, Average or Advanced).

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasContentDepth -->

<owl:DatatypeProperty rdf:about="#hasContentDepth">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
  <rdfs:domain rdf:resource="#RelevanceRelation"/>
  <rdfs:range>
    <rdf:Description>
      <rdf:type rdf:resource="&owl;DataRange"/>
      <owl:oneOf>
        <rdf:Description>
          <rdf:type rdf:resource="&owl;List"/>
          <rdf:first rdf:datatype="&xsd;string">Advanced</rdf:first>
          <rdf:rest>
            <rdf:Description>
              <rdf:type rdf:resource="&owl;List"/>
              <rdf:first rdf:datatype="&xsd;string">Average</rdf:first>
              <rdf:rest>
                <rdf:Description>
                  <rdf:type rdf:resource="&owl;List"/>
                  <rdf:first rdf:datatype="&xsd;string">Basic</rdf:first>
                  <rdf:rest rdf:resource="&rdf;nil"/>
                </rdf:Description>
                </rdf:rest>
              </rdf:Description>
            </rdf:rest>
          </rdf:Description>
        </rdf:rest>
      </rdf:oneOf>
    </rdf:Description>
  </rdfs:range>
</owl:DatatypeProperty>
Figure 3-6: Resource (Fragment) and Concept organization in Concept Ontology
Organization of Concepts and Content in Ontology

<owl:Restriction>
  <owl:onProperty rdf:resource="#isRelevantWith"/>
  <owl:allValuesFrom rdf:resource="#RelevanceRelation"/>
</owl:Restriction>

</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#hasRelevance -->

<owl:ObjectProperty rdf:about="#hasRelevance">
  <rdf:type rdf:resource="&owl;InverseFunctionalProperty"/>
  <rdfs:domain rdf:resource="#Concept"/>
  <rdfs:range rdf:resource="#RelevanceRelation"/>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#toConcept -->

<owl:ObjectProperty rdf:about="#toConcept">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
  <rdfs:range rdf:resource="#Concept"/>
  <rdfs:domain rdf:resource="#RelevanceRelation"/>
  <owl:inverseOf rdf:resource="#hasRelevance"/>
</owl:ObjectProperty>

<!-- http://www.cs.du.ac.in/2011/10/coewr/core.owl#withSearchTopic -->

<owl:ObjectProperty rdf:about="#withSearchTopic">
  <rdf:type rdf:resource="&owl;FunctionalProperty"/>
  <rdfs:range rdf:resource="#Concept"/>
  <rdfs:domain rdf:resource="#RelevanceRelation"/>
</owl:ObjectProperty>

3.4 Organizing Potentially Relevant Web Resources in Concept Ontology

The web resources in the Concept Ontology are organized by forming relevance relation to the relevant concepts. The intent behind this organization is to arrange the web resources that relate concepts and assist a learner by efficiently finding relevant learning resources on a given topic. As mentioned above, URLs of
the topic relevant web resources are stored in CFO based on different resource formats. These resources are linked to concepts in other Domain Ontology. At the same time, these concepts are also linked to web resources as their inverse property. Therefore, this organization enables an application to semantically extract the relevant resources to a given topic.

When a learner enters a query term, the term is expanded semantically depending on the depth of information required by the learner, and then all web resources linked to the expanded terms are retrieved by the system for him/her.

Relevant resources are gathered from the Web by making regular crawls on the concepts that are present in ontology. This task of crawling and gathering web resources can be seen as continuous background task. The resource retrieval and their social semantic ranks, relevant to each of the domain concepts are computed using the algorithm presented in section 3.4.3. The retrieval procedures and computation for relevancy are completed offline before a learner submits her/his query topic. Retrieval of relevant web resources therefore, only requires URLs extraction from ontology. This reduces the effective search time tremendously for a learner. The subtasks carried out by the system to augment the Concept Ontology and managing it are described below.

### 3.4.1 Crawling Potentially Relevant Web Resources

Documents or resources from the World Wide Web and Social sites are gathered using topical crawlers, also called Focused Crawlers. These crawlers find and collect topic relevant web resources which are ranked afterwards according to the relevance. The crawlers can be programmed to work simultaneously on each topic in the Concept Ontology. The working of the crawlers is explained in chapter 4 and the ranking of the potentially relevant crawled resources in chapter 5.

### 3.4.2 Augmenting Concept Ontology

A concept and a URL representing a web resource are linked with each other through a relation. The concept named as ‘dml’ belongs to Domain Ontology and an individual consisting of a URL representing a web resource, belongs to the CFO. An
individual in a content fragment is related to the concepts in the Domain Ontology through the computed Social Semantic Relevance value when queried with the unexpanded query term (original query) using Social Semantic Rank. The domain concept is also linked to the content fragment of CFO consisting of its own relevance to that resource (content fragment) with an inverse property. The Figure 3-7 which shows the number of links to and fro between two concepts need not necessarily is same. URL link concepts in Domain Ontology with various query terms and related content level of web resources; whereas a concept forms inverse link to web resource to represent resource relevance to itself at a particular content level. Therefore every link from a concept to a URL shows the relevance of web resource to that concept with the associated level of content. This simple and clean organization helps to easily extract web resources and present them in an order that is semantically relevant to a concept.

### 3.4.3 Knowledge Management

The background process that is executed automatically at frequent intervals of time to acquire relevant web resources and link them to create knowledge base, is algorithmically summarized below, along with the required set of inputs, complete process and the outputs.
Input:

1. List of \( n \) domains \([D_1, \ldots, D_n]\) that exist in the Knowledge base.
2. Each domain has \( m \) nodes where each node represents a concept, \( C \) under its respective domain, arranged ontologically. Thus we have:

\[
(C_j | C_j \in D_i), \text{ where } 1 \leq i \leq n \text{ and } 1 \leq j \leq m,
\]

where, \( n \) is the number of domains and \( m \) is the number of concepts under their respective domains, (\( m \) may vary for each \( D_i \) as no. of concepts may vary for every subject).

3. Content Fragment Ontology (CFO): consists of the supportive learning material for the underlying domain in the knowledge base. It stores the related URLs and other properties for individual fragment.

Process:

for each \( D_i \):

   for each \( C_j \):

   1. Select \( k \) most semantically relevant URLs to concept, \( C_j \), using Crawling and Ranking Process.

   2. Award Social Semantic Rank (SSR), \( \theta(C_j, URL) \) to each URL

   3. \textbf{for each } \( \theta(C_j, URL) > \delta \)

      ///(where \( \delta \) is the threshold for URL acceptance)

      a. Create URL (resource) instance in CFO based on content format.

      b. Link resource to all concepts of domains contained in expanded query structure, \( Q \) with \( \theta \), \( C_j \) and the expected level of resource content.

   4. Link \( C_j \) to all resources with property as Social Semantic Similarity (\( \theta \)) and level of resource content.
Output:

1. **A set of new URLs (resources) linked to each concept**, that represent web pages consisting of concept related learning material and information.

2. **Automatic updating of knowledge base** with most relevant learning material to domain ontologies.

   The proposed organization of knowledge base and its automatic bindings allow an application to efficiently gather web resources relevant to a given topic. However, this approach requires that the search topic should be present in pre-existing set of Domain Ontology as one of the concepts.

### 3.5 Term Expansion using Concept Ontology

A topic or a given concept has been expanded here using the Concept Ontology. The intend behind the Term Expansion is to reformulate the abstract term (topic) by including its related concepts in the expanded list. This expansion would assist a crawler to find semantically relevant documents on the Web and, rank the crawled documents based on the weighted relevance. The relevance is a unique advantage of the Concept Ontology as was explained in section 3.3. Relevance is an influencing attribute to the process of topic expansion too. For an example, if a learner wishes to have web documents which give a basic idea on a topic of his/ her interest, then the system must able to provide / recommend documents with basic information on the topic and should not recommend documents with in-depth information. On the other hand, another learner who is interested in the depth of the topic may not like to refer that web resource. Instead (s)he would consider a web resource relevant that contains detailed description about the topic. In order to acquire such knowledge, the system needs the expansion and storage accordingly. Therefore, it is very important to expand concepts in a way that the terms in the expanded list may help crawlers to find not only related documents but relevant documents as well.

The Concept Ontology can be used by taking advantage of semantically linked generalized, specialized and related concepts to a given topic for the Topic Expansion. Since one of the motives of the whole work is to retrieve web resources relevant to a learner’s needs, the content level of crawled web resources required by a
A learner/seeker is also required to be taken care of. A learner’s request for the recommended web resources is assumed to be initiated with one of the three options, i.e., the web resources with basic content, average content or advanced level content while submitting a search topic.

Table 3-1: Candidate ‘Terms’ Selection criteria for Topic Expansion

<table>
<thead>
<tr>
<th>Concept nodes from the ontology to be extracted</th>
<th>For Basic Content</th>
<th>For Average Content</th>
<th>For Advanced Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related Concept nodes</td>
<td>None</td>
<td>All direct related concepts</td>
<td>None</td>
</tr>
<tr>
<td>Sub Concept nodes</td>
<td>Direct SubConcepts</td>
<td>Direct SubConcepts</td>
<td>Direct and transitive SubConcepts</td>
</tr>
</tbody>
</table>

The concepts from the ontology are extracted into the expanded list depending on these three information level. The concepts from the ontology are selected in the way so as they can be used to compute the relevance of web pages according to the learner’s requirement. It is believed that if, a learner needs in-depth information on a topic then the concepts that occur down in the hierarchy tree of the search topic which include advanced concepts would help to retrieve relevant web pages.

Table 3-1 lists the concept nodes that are extracted from the ontology for different levels of content requirements. It shows that if the required web resources are expected to contain basic content related to a search topic by a learner, then the expanded list of the search topic must include all super concept nodes (i.e., all concepts that are linked directly or transitively through the hasSuperconcept property from the search topic concept node), and concept nodes linked down the hierarchy from the search topic concept node (i.e., directly linked through the hasSubconcept property). However, for retrieving basic level of content on a search topic, the content nodes linked through isRelatedTo property to the search topic concept nodes do not need to be considered. The reason to this is that isRelatedTo property exists between the sibling nodes in the ontology, and in ontology two terms are placed as sibling when they exhibit different properties.
Therefore, the above argument infers that the sibling concepts do not contribute in providing basic information about a search topic concept.

The sibling nodes are related to the search topic concept, however they may not be semantically relevant to it as meaning of a concept in ontology is described by moving down in the hierarchy. Similarly, the next two columns of the Table 3-1 shows the extraction of the concept nodes from the ontology for the average content level web resources and the advanced level content web resources respectively. In this paper, we are limiting ourselves by focusing only on the retrieval of web resources consisting of basic content information. Based on this expanded list of a search topic, the crawler downloads web pages from a SBS that are tagged with words contained in the expanded list. These downloaded web resources are then computed for the Social Semantic Relevance using the vector space model.

3.6 Discussion

The design of the Concept Ontology was discussed in detail in this chapter. The proposed Concept Ontology was consisting of two different ontologies, Domain Ontology (DO) and Content Fragment Ontology (CFO). The schema of the Concept Ontology was designed to represent DO, CFO and the relationships among their classes. The DO contained the concepts and domains linked through various relations they possess, whereas the CFO consisted of the relevant resources (fragments) linked to the corresponding relevant concept in the DO. The design was created with intent to structurally define the abstract concepts that exist in academic subjects. Thus, the Domain Ontology represents the conceptual knowledge under a domain and, the CFO represents their supportive documents or resources which may help learners to understand the concept of their interests.

The Concept Ontology defines the means in which this knowledge must be used. The Concept Ontology acts as a knowledge base which can be used by many applications for taking decisions and performing intelligent tasks. This includes semantic term expansion, recommending alternate related terms to a learner while searching; crawlers can also use them for gathering web resources. The Concept Ontology being an application independent and curriculum independent design, can
be reused, shared and extended by the Semantic Web enabled applications.

Usage of the Concept Ontology for the Term Expansion was described in the chapter, which is also an important pre-processed input for crawling and ranking of web resources. The work presented in the forthcoming chapters use Concept Ontology as knowledge base.