

# Architecture to Organize Social Semantic Relevant Web Resources in a Knowledgebase

Anjali Thukral, Punam Bedi and Hema Banati

**Abstract**— Learners who believe in self-directed learning usually refer all possible sources to gather information. They also explore social sites to understand the community's view about a topic. At times, this whole process of searching for the required information becomes confusing, time consuming and tedious. Instead of focusing on learning, they tend to spend more time on searching and storing relevant resources suitably in an organized way so that the resources relevant to a concept can be retrieved for later reference. In literature many models have been proposed that have used metadata to organize and store learning content. However they work on a specific customized learning structure intended for a specific institutional or organizational needs. This paper presents an architecture to search relevant resources using a social semantic focused crawler, to organize them in fragment ontology with the term specific relevance and to link each resource to semantically relevant concept in domain ontology, for generic educational use. A prototype of the proposed architecture has been implemented using java, MySQL, Protégé and OWLAPI. The experiment study has been conducted on 'database' domain ontology and fragment ontology. The proposed architecture based content organization provides an easy retrieval of socially semantic relevant resources on the posted query topics.

**Index Terms**—Collaborative bookmarking, eLearning, Social semantic focused crawler, Knowledgebase organization, Ontology, Social semantic relevance.

## I. INTRODUCTION

Internet and WWW are used every day by learners and educators worldwide. They use these web resources for their information need besides the study material and course books or notes. This unbounded information is essentially the need of learners or information seekers. However, they spend lot of time searching resources for the required information and store them in their m/c folders on local disk for later use. Unfortunately, machine and as well as learners, usually, do not remember about the resource contents. Consequently, it results in either repeating the search over the web for fresh retrievals or open each of the stored document to check for relevant content.

The theory of the law of effect given by Edward Thorndike states that: "*When a modifiable connection between a situation and a response is made and is accompanied or followed by a satisfying*

*state of affairs, that connection strength is increased; when made and accompanied or followed by an annoying state of affairs, its strength is decreased"* [1]. This shows that if a learner is provided with the required relevant content at proper time, then it will strengthen his/her behavior towards learning and understanding of the concept. Also, if a learner gets the content which is ranked relevant and popular by the community, a learner will develop more interest in understanding the concept or topic. It may be considered as cognitive influence of social aspect. Therefore, a system that automatically retrieves the topic relevant resources from the web and organizes them in a local knowledgebase for later reference would help a learner to focus more on learning than surfing [2].

This paper proposes an approach to semantically organize web resources in the knowledgebase for learning purpose. These web resources are semantically retrieved from a social bookmarking site and ranked according to the computed relevance to a topic or concept under a given domain. The topic based relevant ranking is computed using domain specific ontology and collaborative annotations tagged by the community. The proposed organization is mainly meant for learning purpose as the ontologies in knowledgebase have been designed and created specifically for concept understanding purpose. However the proposed approach can be utilized for other purposes by incorporating the relevant ontologies.

Remaining paper is organized as follows: Section II briefs research in the area of eLearning, in particular Learning Objects and complexities related to them. The proposed knowledgebase architecture for semantic organization of web resources is presented in section III. The implementation of prototype and benefits of the proposed approach are discussed in section IV following the conclusion and future work in section V.

## II. BACKGROUND

Nowadays, much of the research in the area of eLearning has focused on the organization of learning content so that the multimedia components can be reused that can lead to savings in time and money, and consequently enhance the quality of digital learning experiences [3]. Achieving them requires a well defined organization of information. Before moving towards the proposed work (described in next section), this section firstly reviews Learning Objects and a few of the LOM (Learning Object Metadata) research agendas as proposed by [3], and then discusses some complexities associated with them.

In literature the work has been done on creating or defining the metadata structure for resources that can go well with eLearning environment [4]. Various architectures have been

Manuscript received March 11, 2011. University of Delhi, Delhi, India  
Anjali Thukral is a research scholar with the University of Delhi, Delhi, India. (corresponding author: 9871975544; e-mail: anjali.Thukral@yahoo.co.in).

Punam Bedi is associate professor with the University of Delhi, Delhi, India. She is a senior member of IEEE, a member of AAAI, ACM, and a life member of Computer Society of India. (e-mail: punambedi@ieee.org).

Hema Banati is associate professor with Dyal Singh College, University of Delhi, Delhi, India. (e-mail: hema.banati@gmail.com).

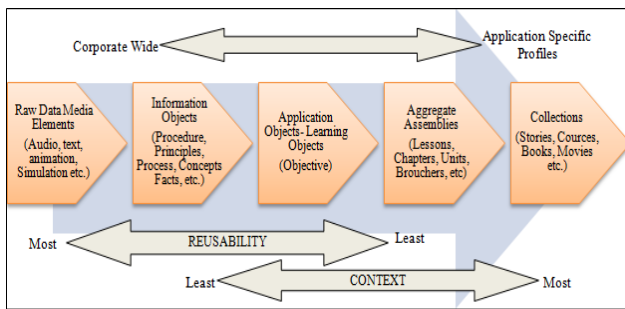


Fig. 1. Modular content hierarchy

proposed which produces satisfactory results but, on the expense of creating new learning material from scratch according to the need of application. SCORM Content Aggregation Model [5] and CISCO RLO/RIO Model [6] are few of the prominent examples of Learning Object Models that identify learning object components and their use. These models by and large generalize the process (Fig. 1) that starts at raw elements and assembles or constitutes these raw elements to information objects that are meant to serve specific objective or problem. The terms like Learning Object, Learning Object Metadata and Learning repositories are frequently used by researchers of this area. A much simplified description of these terms as given by Friesen [7] is as follows. Learning Object refers to digital educational resources; metadata refers to their systematic description to facilitate searching and administration; and repositories represents online, searchable collections of these resources. More specifically, Learning Object Metadata (LOM) standard provided by IEEE Learning Technology Standards Committee defines a Learning Object as: "...any entity, digital or non-digital, that may be used for learning, education or training." [8]. However, this definition allows a wide verity of granularities, which means, Learning Object (LO) to include everything from a few byte content to millions of pages of content.

Duval and Hodgins [3] suggested Learning Object taxonomy which identifies different kinds of LOs and their component to overcome vagueness in the definition. They discussed sixteen Research issues on Learning Objects and their use in education and training. Our paper focuses on the first two research issues among them. The first research issue defines the Learning Object Taxonomy to redefine the basic definition of IEEE LOM Learning Object in view to the vagueness in LO definition. The taxonomy in Fig. 1 shows different levels from domain independent components to more specific application dependent collections. At the most basic level of granularity there exist Raw Data Media Elements, which are contained within the Information Objects. Learning objects contain Information Objects. Aggregate Assemblies contain LOs with some defined objective. This is further assembled into larger collections, like courses and whole curricula.

The second research issue deals with the LO component Architecture and suggests the need to develop a flexible architecture that enables:

- Structuring of LOs and their components that separates content, structure and presentation.
- Interaction between LOs and their components.

These research issues helped in providing a clear vision to work for a quality system. However, there exist some

complexities [7] over the term Learning Object as well. For example it says that the term 'Object' of 'Learning Object' refers to a specific technological paradigm i.e. Object oriented; that consists of specific properties such as polymorphism, concurrency, encapsulation etc. which in contrast, has not been considered by any of the content models. Similarly learning has no clear description as to how the learning occurs or how it can be best understood. Another problem lies with the exchange of content or information among many models that are developed independently with huge investments. Moreover, the architecture requires application based customization. This means one has to put more efforts in developing a complete system for separate institutes, organizations or for other purposes. Even the use of same model with different purpose or institutes requires lot of initial efforts to actually deploy it. Thus, it is hard to provide desired results, unless the whole structure is automatically implemented and semantically linked.

These research issues and objections motivated us to design a structure at the most basic level i.e., at raw data level, but instead of creating new data and objectives, lessons etc., this approach focuses on the reuse of existing voluminous and ever increasing web repositories by automatically annotating them using tags given by the communities and weight them according to their semantic relevance. This will allow a learner to easily retrieve documents that are semantically relevant to their posted query topics.

### III. OUR PROPOSED KNOWLEDGBASE ARCHITECTURE

The proposed architecture of knowledgebase (Fig. 2) consists of ontologies for educational domain. They form semantics among concepts and the available useful web resources. The web resources are kept at the basic level as individuals (instances of fragments). Benefits of ontology over traditional metadata, and in eLearning are explained below following a detailed description of both types of ontology namely, Domain Ontology and Content Fragment Ontology (CFO) in the succeeding subsections.

The intension behind this proposed organization is to arrange the web resources that relate concepts and helps a learner to find relevant learning resources for a given topic efficiently. As mentioned above, URLs of the relevant resources to the topic are stored in CFO structure which is based on various resource formats. These resources are linked to concepts in other domain ontologies. Alternatively, these concepts are also linked to web resources as their inverse property.

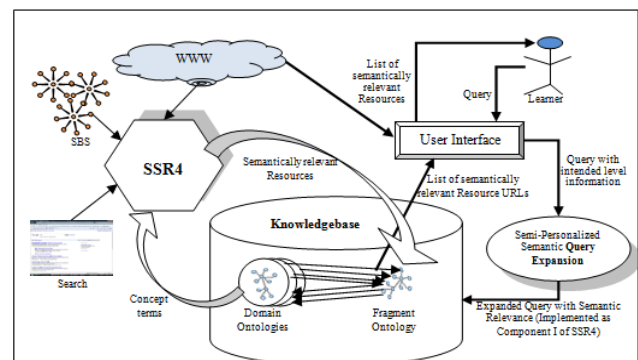


Fig. 2. The architecture of knowledgebase and the related environment.

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 for him/ her.

Crawling and retrieval of the relevant web resources on each concept present in ontology is carried out in frequent intervals of time as a background task. The semantic resource retrieval and their social semantic ranks, relevant to each of the domain concept are computed using an algorithm called Social Semantic Relevant Resource Retrieval and Ranking (SSR4) [9]. Thus retrieval procedures and computation for relevancy are completed offline before a learner submits her/his query topic. Retrieval of relevant web resources on the submitted query by a learner therefore only requires URLs extraction from ontology. This reduces the effective search time tremendously for a learner.

This paper as an extension to [10], particularly focuses on organizing knowledgebase formed by linking concepts in domain ontology and content fragments in CFO. It helps a learner to efficiently retrieve semantically relevant web resources from the knowledgebase. This proposed semantic organization of the web resources ensures to an extent that only relevant resources are delivered to a learner, which in effect helps the learner to save her/his search time.

#### A. Ontology for eLearning Resource Organization

Ontology is a science that studies explicit specifications of the concept and relations among them under a domain [11]. In general philosophical terms ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modeling a domain. The specification takes a form of representational vocabulary (classes, relations, and so forth) which provides meaning for the vocabulary and formal constraints on its coherent use [12]. Moreover, ontology shares a common understanding of the concept organization among community or machines, makes the domain assumptions explicit and enables reuse of domain knowledge. Mathematically, a core Ontology is defined as a structure [13],

$$O := (C, \leq_C, R, \sigma_R, \leq_R, A, \sigma_A, T) \quad (1)$$

consisting of:

- i) four disjoint sets  $C, R, A$  and  $T$  whose elements are called concept identifiers, relation identifiers, attribute identifiers and data types, respectively.
- ii) a partial order  $\leq_C$  on,  $C$  called concept hierarchy or taxonomy,
- iii) a function  $\sigma_R: R \rightarrow C^+$  called relation signature,
- iv) a partial order  $\leq_R$  on,  $R$ , called relation hierarchy, where  $r_1 \leq_R r_2$  implies  $|\sigma_R(r_1)| = |\sigma_R(r_2)|$  and  $\pi_i(\sigma_R(r_1)) = \pi_i(\sigma_R(r_2))$  for each  $1 \leq i \leq |\sigma_R(r_1)|$ , ( $\pi_i(t)$  is the  $i^{\text{th}}$  component of tuple  $t$ , and
- v) a function  $\sigma_A: A \rightarrow C \times T$ , called attribute signature,
- vi) a set  $T$  of datatypes such as strings, integers, etc.

This kind of a graph based hierarchical structure defines a clear association of concepts that simplifies the reasoning tasks among them which makes the ontology suitable for semantically organizing eLearning resources. For eLearning purpose ontologies are desired to:

- define various concepts under different educational domains and

- organize various web repositories that are relevant to the concepts from educational domain.

This provides a separate structure for concepts (topics) and content (learning repositories), consequently easing the process of ontology maintenance and incorporation of frequent changes.

**Advantages of ontology together with collaborative bookmarks over metadata:** In traditional LOM based systems, the content repositories are usually stored in raw form with its properties, and are assembled later for an objective or defined problem. Such architectures store documents with manually annotated metadata and based on their properties different content are assembled to form the Learning Object that are meant to serve some pre-defined objective. Though the metadata related to a document describes it efficiently but have some shortcomings. It firstly requires initial efforts to create metadata manually and secondly, it alone does not fulfill reasoning capabilities as it lacks in semantics [14]. However, Ontology provides a better solution to overcome these issues.

Virtually web resources are huge in number, dynamic in nature and are changing frequently on the web. Feeding metadata manually for each web resource in such a scenario is practically not feasible. Alternatively, there exist social collaborative sites that work as folksonomy (social classification), constructed collaboratively by the communities. Web users on Social Bookmarking Sites (SBS) like delicious.com<sup>1</sup> tag or bookmark web documents with a set of terms (tags) that help them to organize their web resources for later use. The site also allows users to share these bookmarks. Taking advantage of this information as feedback for web resources, their associated bookmarks are used to determine document relevance for a concept through an approach called Focused Crawling using Human Cognition (FCHC) [9]. These retrieved annotations and associated web resources are then applied with pre existing domain ontologies to compute social semantic similarity between the topic and the web resources. Thus instead of metadata, collaborative bookmarks and ontologies are used by this approach to determine semantic relevance of the web resources automatically.

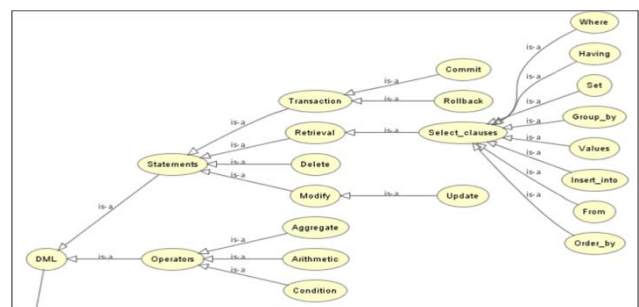


Fig. 3. Snippet of 'Database' domain ontology

#### 1) Domain ontology

Domain ontology in our proposed system consists of a set of ontologies developed manually in protégé<sup>2</sup>. Fig. 3 shows a snippet of 'Database' Domain Ontology. It consists of concepts and relations among them under its own domain. It forms an important part of the knowledgebase and therefore it

<sup>1</sup> <http://delicious.com/>

<sup>2</sup> <http://protege.stanford.edu/>

is created carefully by experts following multiple iterations. Domain ontology that depicts the structure of some topic or domain usually has scope for expansion, but rarely requires any change. Therefore, domain ontology needs only one time effort per domain.

## 2) Fragment ontology

The web resources are stored in fragment ontology as instances of the content fragments which are based on Content Fragment Classification [15]. The intention behind using this classification is to separate the structure from content which is an important step towards reusability, also suggested by [3] in their Research agenda 2. Therefore, the content fragment is used as the most basic form of content that represents individual resources.

In addition to it, this approach also stores the associated social semantic relevance with concepts and the level of content which is computed for each resource. Resource annotation using this information is processed automatically which is an advantage over the manually created metadata. This annotation helps to retrieve relevant resources from ontology for learners with effectively reduced computational time. URLs of semantically relevant web resources to the domain concepts are organized according to their format in the ontology which is named as **Content Fragment Ontology (CFO)**.

Each fragment in CFO is linked to the concept(s) along with the computed relevance with respect to some concepts of Domain ontology. CFO makes an early distinction between classes that describes discrete elements and continuous elements. 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. Fig. 4 illustrates CFO classification.

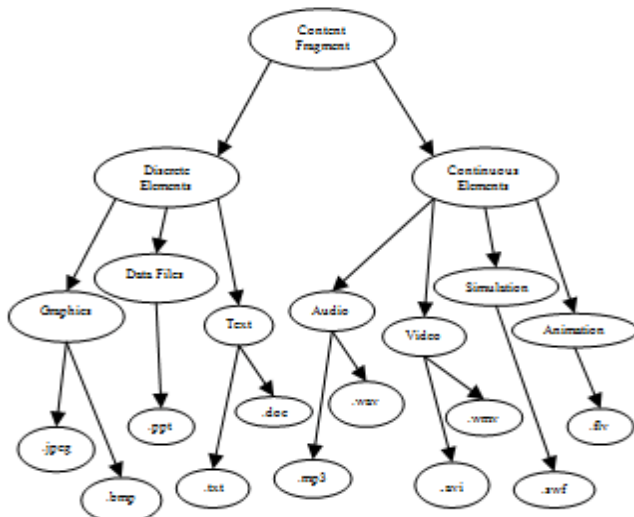


Fig. 4. Content fragment ontology based on content fragmentation classification

## B. SSR4: Social Semantic Relevance Ranking of Retrieved eResources

The SSR4 module is composed of three parts: The first part is semantic query expansion which is used to expand the entered query using domain ontology from the existing

knowledgebase. The process of expansion is carried out by considering basic, average and advanced information levels of the content. This component thus, computes the Semantic Relevance for each of the expanded term and calculates the query vector which is later used by part II and III. The second part is responsible for searching relevant eResources by taking seed URLs from a search engine's search results. Starting with these seeds, it crawls into a SBS using a focused pattern search called FCHC, which searches semantically relevant web resources tagged by SBS users. The third part takes the query vector which is computed using Vector Space Model based on ontology (from the first part) and semantically relevant eResources extracted from a SBS (from the second part) as input. The social semantic relevance ranking is then computed using cosine similarity between the resource vectors and the query vector.

### 1) Semantic query expansion

The Semantic Query Expansion (SQE) component uses a knowledge base that consists of multiple domain ontologies, constructed manually for education purpose. These ontologies represent hierarchical taxonomies consisting of concepts and relations among them. It extracts semantic relevant terms from domain ontology to expand the posted query term or topic. The expansion is carried out, based on the content level required by the learner (personalization). The system provides three options to a learner: content relevant at *basic level* for a beginner, *average level* for related articles or *advanced level* for in-depth information on a given topic. The query vector length  $|Q|$  is computed using the semantic distance between the query term and its each expanded term as the following function:

$$|Q| = \sqrt{\sum_{q_t} Wt_{Q_o, q_t}^2} \quad (2)$$

Here,  $Wt_{Q_o, q_t}$  is the weight assigned to each  $q_t \in Q$ ,  $\forall (n1 + 1)$  terms, It is formulated as following:

$$Wt_{Q_o, q_t} = \frac{SR_{Q_o, q_t}}{\sum_{t=0}^{n1} SR_{Q_o, q_t}} \quad (3)$$

SR is the Semantic Relevance computed using the domain ontology that contains the query topic.

### 2) Social semantic focused crawler

A social semantic focused crawler FCHC (Focused Crawling using Human Cognition) [9] has been used to retrieve relevant resources. The pattern used to search relevant resources in social bookmarking site (SBS) makes this crawler different from others. It explores the pages of SBS users for bookmarked web resources with tags that are related to the semantically relevant terms.

### 3) Social semantic resource ranking

The Semantic Relevance of tags is computed for all SBS bookmarks which are semantically relevant. The social semantic relevance of each tag associated with each resource is computed by incorporating the popularity of the resource among SBS users. The resource vector length is then calculated for each resource using the tags weight computed by the social semantic relevance. Following formula is used to calculate the resource vector length.

$$|r_i| = \sqrt{\sum_k Wt_{t_k, r_i}^2} \quad (4)$$

The distance is measured between the query vector and

each resource vector using the cosine similarity [16].

$$\theta_{Q_o,r_i} = \frac{\sum_{q_t=t_k}(Wt_{Q_o,q_t} \cdot Wt_{t_k,r_i})}{\sqrt{\sum_{q_t} Wt_{Q_o,q_t}^2} \sqrt{\sum_{k} Wt_{t_k,r_i}^2}} \quad (5)$$

These resources are then ranked by sorting them on the computed  $\theta$  in descending order. The value of similarity measure ( $\theta$ ) is maximum for the most relevant resource. The relevance of a resource to the query term decreases with the decreasing  $\theta$ .

### C. Linking semantically relevant web resources in Knowledgebase

The classes and subclasses of CFO are called Content Fragments whereas, the classes and subclasses of Domain Ontology are called concepts, in order to distinguish the objectives of both type of ontologies. Relevant web resources that are retrieved from WWW using SSR4, forms the instance (individual) of content fragments. Links in the knowledgebase are established between these instances and concepts in domain ontology. An overall view of both ontologies and the relevant resources showing their linking with each other is illustrated in Fig. 5. It shows two ontologies: domain ontology consisting of concepts and content fragment ontology consisting of content fragments.

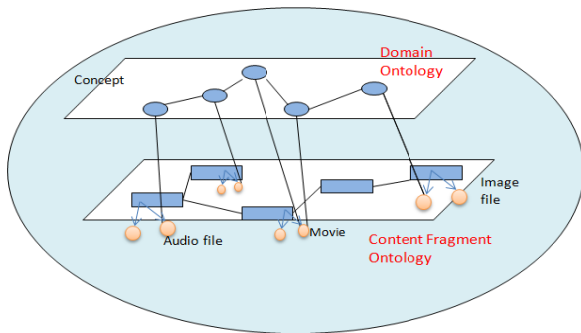


Fig. 5. Resource organization in the knowledgebase

These content fragments have instances in the form of URLs. These URLs are the web resources that have been retrieved from SSR4 using some query topic. Therefore, each resource in turn is associated with a set of concepts in domain ontology. Using this information, the resources which are instances of content fragment are linked to the associated concepts in domain ontology. These links or relations also store additional information like the relevance score of a resource to a concept and the level of content in the resource.

A concept and a URL representing a web resource are linked with each other through a relation is illustrated in Fig. 6 as a specific case. 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 concept of domain ontology through the computed social semantic relevance value when queried with the unexpanded query term (original query) using Social Semantic Rank (SSR) algorithm [9]. 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 example in Fig. 6 clearly shows that number of links to and fro between two concepts need not necessarily be 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 a 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.

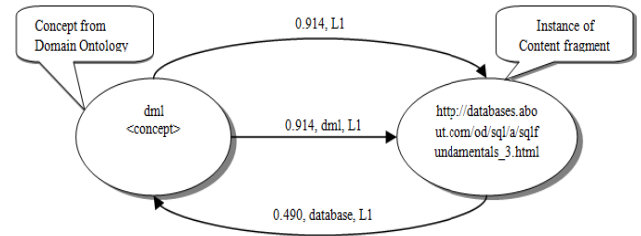


Fig. 6. Links between a web resource and a concept from domain ontology

## IV. IMPLEMENTATION

A prototype is implemented with domain ontology and a CFO using java, MySQL and OWLAPI. SSR4 uses a social semantic focused crawler FCHC that retrieves relevant web resources to feed the knowledgebase. The number of relevant resources is taken above a particular threshold because resources with a small relevance are usually of no interest for the learner. The table-1 shows the top 15 resources retrieved by the FCHC focused crawler and the similarity computed by SSR4.

The Domain ontology and CFO are built manually in Ontology editor Protégé, whereas, web resources that are retrieved using SSR4 algorithm are individually annotated automatically. Besides annotations, semantic links are created among concepts in domain ontology and individuals (instances/ web resources) of content fragments. The background process that is executed automatically at frequent intervals of time to acquire relevant web resources and link them to create knowledgebase is algorithmically summarized below, along with the required set of inputs, process (as illustrated in Fig. 2) and the outputs.

TABLE I: SOCIAL SEMANTIC RELEVANT ERESOURCES RETRIEVED BY SSR4

Relevant URLs to the concept 'DML'	Cosine Similarity	Rank
<a href="http://databases.abour.com/od/sql/a/sqlfundamentals_3.htm">http://databases.abour.com/od/sql/a/sqlfundamentals_3.htm</a>	0.914236047	1
<a href="http://en.wikipedia.org/wiki/Data_Manipulation_Language">http://en.wikipedia.org/wiki/Data_Manipulation_Language</a>	0.907737623	2
<a href="http://www.tomjewett.com/dbdesign/dbdesign.php?page=ddl_dml.php">http://www.tomjewett.com/dbdesign/dbdesign.php?page=ddl_dml.php</a>	0.900403891	3
<a href="http://www.informit.com/articles/article.aspx?p=32101">http://www.informit.com/articles/article.aspx?p=32101</a>	0.892203025	4
<a href="http://www.sqlservercentral.com/articles/Audit/65585/">http://www.sqlservercentral.com/articles/Audit/65585/</a>	0.868223891	5
<a href="http://www.relationalsolutions.com/visHelp/WebHelp/Excel_Object.htm">http://www.relationalsolutions.com/visHelp/WebHelp/Excel_Object.htm</a>	0.846418108	6
<a href="http://www.oraFAQ.com/faq/what_are_the_difference_between_ddl_dml_and_dcl_commands">http://www.oraFAQ.com/faq/what_are_the_difference_between_ddl_dml_and_dcl_commands</a>	0.82814531	7
<a href="http://www.c-sharpcorner.com/UploadFile/skumaar_mca/XMLDmlOperation07292009123939PM/XMLDmlOperation.aspx">http://www.c-sharpcorner.com/UploadFile/skumaar_mca/XMLDmlOperation07292009123939PM/XMLDmlOperation.aspx</a>	0.822570545	8
<a href="http://dotnetguts.blogspot.com/2007/06/transaction-control-language-tcl.html">http://dotnetguts.blogspot.com/2007/06/transaction-control-language-tcl.html</a>	0.822570545	9
<a href="http://ora.u440.com/dml/update.html">http://ora.u440.com/dml/update.html</a>	0.81381437	10
<a href="http://technet.microsoft.com/en-us/library/cc280358.aspx">http://technet.microsoft.com/en-us/library/cc280358.aspx</a>	0.802982722	11
<a href="http://msdn.microsoft.com/en-us/library/aa140015(office.10).aspx">http://msdn.microsoft.com/en-us/library/aa140015(office.10).aspx</a>	0.799044264	12
<a href="http://esbudellat.net/dml/1.0/">http://esbudellat.net/dml/1.0/</a>	0.798010645	13
<a href="http://www2.aao.gov.au/2dFGRS/Public/Release/Database/sql_intro.pdf">http://www2.aao.gov.au/2dFGRS/Public/Release/Database/sql_intro.pdf</a>	0.798010645	14
<a href="http://www.cs.sfu.ca/CC/354/zaiane/material/notes/Chapter1/node17.html">http://www.cs.sfu.ca/CC/354/zaiane/material/notes/Chapter1/node17.html</a>	0.798010645	15

**Input:**

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

$$(C_j | C_j \in D_i),$$

where  $1 \leq i \leq n$  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 knowledgebase. 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 SSR4.
2. Award Social Semantic Rank (**SSR**),  $\theta(C_j, URL)$  to each URL, again computed using SSR4.
3. for each  $\theta(C_j, URL) > \delta$  // (where  $\delta$  is the threshold for URL acceptance)
  - i) **Create URL (resource) instance** in CFO based on content format.
  - ii) **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 focused crawler, FCHC executes at frequent intervals of time by the module SSR4 on all concepts of domain ontology to update the knowledgebase. All links among the concepts and instances of content fragments are established to incorporate latest changes.

Thus, the proposed organization of knowledgebase and its automatic bindings allow a learner to efficiently retrieve web resources relevant to a given topic. This approach assumes here that these topics that are searched by learner also exist as one of the concepts in pre existing set of domain ontology.

The results obtained from the above algorithm are shown as a list of URLs to user that can be opened in a browser after getting connected to internet. One of the important benefits of the proposed system is that, the retrieved resources are assessed semantically and socially by the community and therefore retrieve relevant resources only. Iterative execution of retrieval and ranking process saves lot of time of learner as it completes maximum of searching and ranking process well before the query submission. This provides learner with

instant results that helps in enhancing the quality of digital learning experiences.

V. CONCLUSION

The architecture to organize social semantic relevant web resources in a knowledgebase has been presented in this paper. The knowledgebase is organized using two types of ontologies, which separate the concept structure from content or resources (web documents, audio files, videos, text documents, presentations etc.) at the basic levels. They are domain ontology and content fragment ontology (CFO). The domain ontology consists of various semantically linked concepts that represents individual domain for educational or learning purpose; whereas CFO classifies various web resources as content fragments for structuring.

The topic relevant web-resources are retrieved by the module SSR4 (Social Semantic Relevant Resources Retrieval and Ranking) which uses a social semantic focused crawler FCHC. It crawls into a social bookmarking site by analyzing tags given to the resources in a systematic search pattern. The retrieved resources form the individual of content fragments and they are also semantically linked to the concepts (from domain ontology). These links also carry the similarity value between the query term and resource when searched for different level of content (depth) by a learner.

A prototype has been implemented with 'Database' domain ontology and CFO. A learner is allowed to retrieve documents on any topic (query) under 'Database' domain, through a user interface. The query is semantically expanded with (optional) additional information which involves the learner's requirement for the level (depth) of resource content, using part I of SSR4. Based on this expanded query, the relevant resources above a threshold, which is fixed by the learner, are retrieved from the knowledgebase. The resource retrieval for a learner does not include the time spent in searching and retrieving documents from the WWW and the SBS as this done offline. Therefore, the approach reduces the effective searching time for a learner tremendously, which consequently, enhances the quality of digital learning experience. A learner needs to be online only when he/ she wish to open a web page using the recommended URLs.

The work is in the process of integrating all modules (SSR4, query expansion, FCHC and knowledgebase organization), which have presently been implemented separately.

REFERENCES

- [1] J. Bolton, "Leraning theories in education," Making waves in education.: University of York, pp. 1-17, 2008.
- [2] P. Bedi, H. Banati, and A. Thukral, "Use of Ontology for Reusing Web Repositories for eLearning," Technological Developments in Networking, Education and Automation, K. Elleithy, T. Sobh, M. Iskander, V. Kapila, M. A. Karim and A. Mahmood, eds., Netherlands: Springer Netherlands, pp. 97-101, 2010.
- [3] E. Duval and W. Hodgins, "A LOM Research Agenda," Proc WWW2003 Conference, 20-24 May, Budapest, Hungary, 2003.
- [4] A. Thukral, R. Sharma, H. Banati, P. Bedi, "Modeling Roles in E-learning: MoRel," Proc. International Conference on Advances in Computer Vision and Information Technology, pp. 612-620, 2007.
- [5] Sharable Content Object Reference Model (SCORM®) 2004 4th Edition Content Aggregation Model (CAM) Version 1.1.: Advanced Distributed Learning (ADL), 2009.

- [6] C. Barrit, D. Lewis and W. Wieseler, "CISCO Systems Reusable Information Object Strategy," Version 3.0. USA : Cisco Systems, Inc, 1999.
- [7] N. Friesen, "Three Objections to Learning Objects and E-Learning Standards," Online Education Using Learning Objects, R. McGreal (Ed.), pp. 59-70, available at <http://learningspaces.org/n/papers/objections.html>, 2004.
- [8] LOM, IEEE LTSC. Draft Standard for Learning Object Metadata IEEE 1484.12.1-2002. New York, USA : IEEE 1484.12.1-2002, 2002.
- [9] P. Bedi, H. Banati and A. Thukral, "Social Semantic Retrieval and Ranking of eResources," Proc. Second International Conference on Advances in Recent Technologies in Communication and Computing ACEEE, 2010, Kerala, India : pp. 343 - 347, 2010.
- [10] A. Thukral, P. Bedi, and H. Banati, "Automatic Organization of Web Resources in Ontologies for Learning Purpose," Proc. 2nd International Conference on e-Education, e-Business, e-Management and E-Learning (IC4E 2011). pp. 38-44. Mumbai, India. : IEEE, 2011.
- [11] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition, 5(2), pp.:199-220, 1993.
- [12] T. Gruber, Ontology (Computer Science) - definition in Encyclopedia of Database Systems. Ling Liu and Tamer M Özsu. [ed.], Encyclopedia of Database System, 2008.
- [13] P. Cimiano, Ontology Learning and Population from Text: Algorithms, Evaluation and Applications, Springer-Verlag New York: Springer, 2006.
- [14] X. Wang, F. Fang and L. Fan, "Ontology-Based Description of Learning Object," Proc. 7th international conference on Advances in Web Based Learning table of contents, Jinhua, China : Springer-Verlag Berlin, Heidelberg , pp. 468 - 476, 2008.
- [15] K. Verbert and E. Duval, "Towards a global component architecture for learning objects: An ontology based approach," Proc. Workshop on Ontologies, Semantics and E-learning., Agia Napa, Cyprus, OTM 2004 pp. 713-722, 2004.
- [16] E. Garcia, Cosine Similarity and Term Weight. Information Retrieval Tutorials. [Online]. Available: <http://www.miislita.com/information-retrieval-tutorial/cosine-similarity-tutorial.html>, 2006.

**Anjali Thukral** is an assistant professor in the Department of Computer Science at Keshav Mahavidyalaya, University of Delhi with 10+ years of teaching experience. She is also a research scholar and her research interests include eLearning, Information Retrieval, Knowledge Representation, Semantic Web and Multi-agent Systems.

**Dr. Punam Bedi** received her Ph.D. in Computer Science from the Department of Computer Science, University of Delhi, India in 1999 and her M.Tech. in Computer Science from IIT Delhi, India in 1986. She is an Associate Professor in the Department of Computer Science, University of Delhi. She has about 25 years of teaching and research experience, and has published about 90 research papers in National/International Journals/Conferences. Dr. Bedi is a senior member of IEEE, a member of AAAI, ACM, and a life member of Computer Society of India. Her research interests include eLearning, Semantic Web, Information Retrieval, Web Intelligence, Intelligent Information Systems, Soft Computing, Multi-agent Systems, Trust, Personalization, Intelligent Software Engineering, Intelligent User Interfaces, Requirement Engineering and Human Computer Interaction (HCI).

**Dr. Hema Banati** is an associate professor in the Department of Computer Science in Dyal Singh College, University of Delhi. She received her Ph.D. in Computer Science from the Department of Computer Science, University of Delhi, India in 2006. Her research interests include eLearning, Web Engineering, Software Engineering and Human Computer Interfaces and Multi-agent Systems.