CHAPTER 1

Introduction

- Introduction
- Problem
- Need for the present work
- Objectives
- Scope of the Work
- Literature Review
- Hypotheses
- Methodology
- Chapterization
- Conclusion
- References
1.1 Introduction
The web is a success story both in terms of availability of information and in terms of increasing number of users. Today people use the web for different purposes, including knowledge acquisition, sharing thoughts with others, business purposes, entertainment among other topics. However we all are aware about the vagueness and noise factor in retrieved sets from the Web using the present search engines, like, Google, Yahoo, etc). Also, it is true that today most of the data on the Web are for human consumption and for human interpretation, not for machine interpretation.

Web service is a software system designed to support interoperable machine-to-machine interaction over a network [1]. Web services are the procedural extension of the existing World Wide Web (WWW). They basically turn the distributed set of information of the web pages into a distributed set of services. But at present, web services are struggling to expand against the limitations of existing web architecture. There also exist conflicts with the proprietary standards.

Semantic Web [2], an extension of the present web is characterized by the association of machine-accessible formal semantics with more traditional Web content. The motivation behind Semantic Web is to increase automation in processing Web-based information and to improve interoperability of Web based information systems. [3]. Semantic Web services [4] are semantic extensions of the existing web services which support the automatic discovery, composition, invocation and interoperation of services. Researchers both in academia and industry are investigating how Semantic Web services can be described and what kind of operational environment is needed.

1.2 Problem
Information available on the Web is encoded mostly using the HyperText Markup Language (HTML). But the information available within HTML is difficult to manipulate on a large scale. This is true even if the
Chapter 1: Introduction

representation is done using eXtensible Markup Language (XML) [5] due to large variations of tags and vocabulary used. XML basically provides us a way to organize data and documents in a structured and universally accepted format using tags. However, these tags do not have any predefined meaning [26].

The keyword based search engines do not have the capacity to understand users query and retrieve the information by making a semantic matching and at present it is mostly done by just syntactic matching only. Computers are not in a position to process and infer any knowledge from the stored information. Human beings use a lot of background knowledge and intuition to interpret and use the data on web. Consider eLearning scenario, where a learner with a particular educational need, visits one or several online educational centers; browse their resources; collect information about the courses, such as, study programs, prerequisites, needed software tools, fees, etc.; then select the most appropriate course for his/her needs and preferences and, finally, register for the selected course. Such information is generally scattered in different pages in different educational establishments’ pages. So a learner needs to collect all these information manually. This manual browsing is too time consuming and, most importantly a user has to visit a few online centers before making a decision.

The present day demands are [23],

1. A software program must be able to automatically find, or discover, an appropriate service.
2. Software must be able to determine automatically how to invoke or execute the service.
3. Software must be able to select and combine a number of web services to complete a certain objective. So, the services have to interoperable with each other seamlessly to provide a combined result and valid solution.
4. The software must be able to verify and monitor the service properties while in operation.
Chapter 1: Introduction

Whereas, the present day web technology fails to fulfill the above mentioned demands, there is potentiality that using Semantic Web technology it would be possible to achieve those goals.

### 1.3 Need for the present work

The main problem with HTML language is that it is basically made for data presentation on the web, not for data interpretation by machine. So, what we need is that data should be described in such a way that different software agents can interpret them and can process and generate automatic semantic web services. For this it is most important to use rich and standard language to describe domain knowledge. But only rich and standard language alone cannot solve the problem. When we expect the machine to deliver more personal and/or customized information, we need to encode the well-grained data. The present study is an effort towards the structuring and description of data in a machine friendly manner and to give emphasis on fine-grained data. The study also concentrates on binding the context, so the system delivers meaningful information according to user needs in a particular context.

In order to meet the present day demand that a software agent must be able to automatically discover, compose, invoke or execute the web services, we need to describe the services in a machine interpretable way. In this regard we concentrate on describing the services using a standard and rich language. To demonstrate semantic matching and retrieval, a need also arises to develop a model for generating automatic semantic web services.

### 1.4 Objectives

The objective of the present work is to study the feasibility of the Semantic Web technologies in improving information retrieval on the Web. In addition, an attempt will be made to develop a generic model, which will be applicable across the domains.
The main objectives of the present study are:

1. Description of domain knowledge in machine processable form using ontology language.
2. Description of the services semantically.
3. Derive a common model in generating semantic web services across the domains and also develop techniques for interaction among different semantic web services.

1.5 Scope of the Work

The idea of semantic approach is to express information in a machine processable way. In this regard, Semantic Web technology has drawn a considerable attention of the researchers in the field of distributed information systems, artificial intelligence, and so on. Researchers in the distributed information systems are taking interest to make use of Semantic Web technology as a central component of their software constructions [6].

The methodology involved application of Semantic Web technologies to achieve the goal of context based retrieval. There are many new tools and techniques, standards that have emerged. Considering vastness and diversity of the Web (e.g., its vastness, multi-dimensionality), and also the emerging nature of Semantic Web technologies, the scope of the present work encompasses the areas as explained below:

1. Extension and application of the Semantic Web Technologies (SWT) in general. Since the goal of the present work is to build a system based upon the emerging Semantic Web technologies, it is important to take understand and trace the developments in SWT in general and moreover trace SWT applications in particular to information organization, services and transactions.

2. Modeling for generating web based information services which will be applicable across the domains of knowledge and will be able to semantically associate concepts, organizations, services and such.
Chapter 1: Introduction

For example, eLearning, eGovernance, eHealth, eCommerce, etc. For the purpose of demonstration of the system in practical environment, we deploy the services in arena of eLearning which we treat as 'application domain'.

3. The main associated area of work will be in 'Ontologies'. The area of ontologies is much discussed in academic and research circles. Different methodologies and approaches have been adopted for building ontologies. The present work will have a focus in the area of ontologies and ontology building methods. Various distinct applications will studied and with particular emphasis on applications of ontologies in information services. Since the work involves practical demonstration it is intended to develop the ontologies. The application area of our system is eLearning domain, the different distinct 'facets' or components of eLearning are traced. Accordingly the ontology architecture will be modular comprising modules for domain ontology, document ontology and student ontology.

4. Ultimately we examine the application of SWT and the work will encompass semantic approach to generating services and methods in encoding for such services.

1.6 Literature Review

We conducted our literature search with emphasis on the Semantic Web and eLearning areas. Since we considered eLearning as our application domain, it is important for us to understand the present state-of-the-art of both related domains viz., Semantic Web and eLearning. There are several significant efforts as evidenced in literature that researchers in either or both of the mentioned areas, tried to apply the new emerging technology to improve the eLearning scenario and to improve the learning experience of the learners. The following sections trace the developments and some findings in the above mentioned related area of study. The literature search
has consulted Internet resources, the online bibliographic and full text material, journals, reports and books.

1.6.1 Source of Literature
Following are some of the important sources consulted in surveying the literature.

LISA: Library and Information Science Abstracts is an international abstracting and indexing tool designed for library professionals and other information specialists. LISA currently abstracts over 440 periodicals from more than 68 countries and in more than 20 different languages. Over total 322,221 records are available as of September 2009 (coverage is starting from 1969 to current) [7].

The ACM Digital Library: The ACM DL is a vast collection of citations and full text from ACM journal and newsletter articles and conference proceedings [8].

IEEE Explore Digital Library: a full text database, providing access to the world's highest quality technical literature in electrical engineering, computer science, and electronics [9]

Science Direct: ScienceDirect is a leading full-text scientific database offers articles/chapters from more than 2,500 peer-reviewed journals and more than 11,000 books. There are currently more than 9.5 million articles/chapters. Since 2003, it started hosting value-added content, such as, audio and video files, datasets and other supplementary content associated with the research [10].

Emerald: It is a major database giving access to over 200 journals and a wide range of serials, series and books in management and related fields [11].
Springer Link: It is one of the world's leading interactive databases for high-quality STM journals, book series, books, reference works and the Online Archives Collection. [12]

CiteSeerX: It is a scientific and scholarly literature digital library and search engine that focuses primarily on the literature in computer and information science. [13]

Google Scholar: It is a freely-accessible Web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. It allows users to search for digital or physical copies of articles, whether online or in libraries [14].

Added to the above sources, we also used the regular search engines (e.g., regular Google search engine, Yahoo, etc.) in conducting our literature survey. In addition journals, reports and books as refereed in the sections below also threw light upon the developments and landmarks in the related areas of this study.

1.6.2 Related Works
It is often deemed that the research work on the Semantic web and eLearning has not been moved significantly forward as per the expectations. But still there are certain significant number of experiments/ attempts in research projects, experiments, models and systems. It is important to analyse those and resolve the open issues. Here we discuss some of the significant studies.

Stojanvoc, L., et al. [15] applied semantic web technology to implement an eLearning scenario. Their work is primarily based on ontology-based descriptions of content, context and structure of learning materials. The work focused on dynamically creating course structures and not on the mediating of learning services. They used F-Logic in representing their ontology.
Project Elena [16] focused on the combination of semantic web and p2p technologies. It is service-oriented perspective, which defines a smart learning space of educational service providers based on the Edutella [17] peer-to-peer framework for interoperability and resource ex-change between heterogeneous educational applications and different types of learning resource repositories. Smart Spaces for Learning offer intelligent services for the users. Personalisation techniques support learners in their searching for learning resources. The 'intelligence' of their tool is based on sophisticated learner profiles and ranking patterns according to preferences, interests, learning goals and learning history.

Henze, N., et al. [18] tried to resolve the most significant challenge of Semantic Web, the provision of distributed information with well-defined meaning, understandable for different parties. Particularly, concentrated on, applications that are able to provide individually optimized access to information by taking the individual needs and requirements of the users into account. They envisioned personal learning services capable of interpreting metadata-annotated learning resource, understanding their annotations with respect to standard ontologies for learning materials, like e.g. LOM or IMS. They tried to show how the semantic web resource description formats can be utilized for automatic generation of hypertext structures from distributed metadata. They investigated ontologies and metadata for different types of resources such as, user and observation. Their work carried a logic-based approach to educational hypermedia using TRIPLE, a rule and query language for the semantic web.

Santos, J. M., et al. [19] applied semantic web technology to design brokerage architecture for eLearning domain. They created several sub-ontologies, for instance, courses and learning objects, on-line service providers, content developers, learners, etc. They used descriptive logic (OWL-DL) for developing ontologies. RDQL is used for expressing the queries.
Ghaleb, Fayed F. M., et al. [20] provides a framework for web-based eLearning system using the Semantic Web technology. Presented an approach for implementing a Semantic Web-based eLearning system, which focus on the RDF data model, OWL ontology language and RAP (RDF API for PHP, a Semantic Web toolkit for PHP developers) for parsing RDF documents. The Protégé 2000 ontology editor is used to create the eLearning ontology classes and properties. The significance of the proposed model is that it contains a hierarchical contents structure and semantic relationships between concepts which can provide related useful information for searching and sequencing learning resources in web-based eLearning systems.

LUISA [21] is a STREP Project under the European Commission's 6th Framework Programme. The outcome of the LUISA project is the development of a Semantic Web Service-based Architecture for Learning Object Discovery, Selection, Negotiation and Composition within a distributed service architectures seamlessly integrated through ontologies. It applied the architecture and tools developed with the DIP project, including IRS-III, to eLearning. More specifically, LUISA focus is to create a rich flexible infrastructure supporting the development and reuse of learning materials for both learners and educators.

Fiaidhi, J., et al. [22] worked on a Multimedia Learning Object Brokering Architecture implemented using Apache Axis, Jena and Pellet. The Broker component determines which Multimedia Learning Objects satisfy a query based upon information contained in one or more domain-specific ontologies. The system’s requestor and provider components are designed specifically for use with SVG slideshow presentations described using the CanCore standard. They used Descriptive Logic (OWL-DL) as the representation for their ontology. All queries are expressed in OWL-OQL (OWL- Object Query Language) (not recommended by W3C), a concise OWL query language created for use with their brokering system.
Chapter 1: Introduction

1.7 Hypotheses

The present study is based on the hypotheses,

- Domain knowledge base can be developed using ontology language
- The web services can be described using ontology languages for web services;
- Semantic web services have cross-domain applications.

1.8 Methodology

One of the primary goals of the present study was to build a system based upon the emerging Semantic Web technologies (SWT) in order to achieve context based information retrieval and services. So, it was important to understand and trace the developments in SWT in general and moreover trace SWT applications in particular to information organization, services and transactions. In order to do so, we first made an extensive study of the present Semantic Web technologies and tools. We prepared a state-of-the-art-report for the emerging Semantic Web technologies, tools and techniques. Following the investigation on technology, tools and techniques, our immediate goal was to choose an application area to demonstrate our system efficiency and effectiveness. We chose eLearning as an application domain. The next course of action was to understand eLearning and its characteristics; and the learners’ characteristics and their requirements. For that, we made an extensive literature survey and a report was prepared.

The next step was to identify the issues involved in the present eLearning systems. So, we carried out an extensive literature survey and enlisted the issues based upon their immediate importance in the present scenario. We translated those existing eLearning system issues, including the learners’ characteristics into a list of system competencies emphasizing the student perspectives.

Metadata, an Internet-age term which librarians traditionally have used to classify books and other printed documents, also plays crucial role in our
present work. It is because, metadata, at its most basic level, provides a common set of tags that can be applied to any resource, regardless of who created it, what tools they used, or where it is stored. For that reason we made an extensive study on the existing metadata standards both for document description and learner description. And then depending upon our requirements and considering the most popular standard, we selected the standards to be used in our present work.

The next step was to identify the standard in representing the domain knowledge and primarily we selected SKOS, Simple Knowledge Organization System and further extended it to achieve our goal. Then we prepared an application profile for the used properties (roles and attributes) along with their domain and range, inverse property, logical characteristics and cardinality constraints in domain ontology. Our next step was to design a conceptual model (component architecture), a Semantic Learning Layer Cake which will be applicable to any other domains, as the model is developed on the most basic concepts, such as, content, context and structure. We then developed modular ontologies, such as, domain ontology, document ontology, and learners’ ontology. We chose OWL-DL in developing the ontologies and application profiles enlist the used classes, semantic properties, and constraints in the ontologies.

The next step was to develop the algorithms to generate the services and encoded the services semantically. This had been done using a semantically rich language, N3Logic. The next step was to instantiate the knowledge-base. For that purpose, we considered indexing and abstracting as a learning course under the programme Library and Information Science. In this regard we developed the domain ontology (methodology is described in chapter 5) for the indexing and abstracting domain. The document ontology is populated using the lessons for the course indexing and abstracting. We also populated our learners’ ontology base by adding some sample data of learners. Our next assignment was to test the system against the system competency. In this regard we demonstrated some use cases against the enlisted system competencies. At the end we enlisted our research findings
and observations, system limitations and further research questions in the conclusion chapter.

1.9 Chapterization
The present study is organized into the following chapters.

1. Introduction
This is the introductory chapter. It introduces goal and objectives of the study including the need of the present study, hypothesis, scope, related works being carried out by other researchers and research methodology.

2. eLearning and Learning Management Software Tools
This chapter describes the eLearning, learning objects, characteristics of eLearning and eLearning benefits, characteristics of the learners, open source software and learning management systems and evolution of eLearning.

3. Learning Standards
This chapter deals with the learning standards, emphasizing the metadata standards used in elearning.

4. Semantic Web Technologies
This chapter discusses the Semantic Web and Semantic Web technologies, including the ontology, ontology languages and semantic web services.

5. eLearning Services Using Semantic Web Technologies
This is the main system chapter. It deals with the verification of the present eLearning system and the present relevant issues within it, conceptual framework of the semantic eLearning, system architecture, ontology frameworks and ontologies, algorithms and the service encoding.

6. Conclusion
This is the concluding chapter. It provides a summary of the entire study discussed in different chapters, and notes down the overall observations and suggestions based upon study made in different
Chapter 1: Introduction

chapters. It also discusses some system limitations and further research questions.

1.10 Conclusion

The knowledge representation techniques of semantic web technology in a distributed environment have tremendous potential applicable for richer description of web resources. It is demonstrated that using semantic web technology, we could enhance web based information services.

eLearning domain should essentially cater to learner based information requirements. Accordingly eLearning environment becomes unique in the sense that the whole learning cycle is transferred to the online mode not only the lessons and learning materials. In this case an eLearning system is expected to be highly intuitive. The present work has demonstrated the use of Semantic Web technologies in providing “meaningful services” to the learners and demonstrates the same in the application domain of eLearning.

1.11 References


Chapter 1: Introduction

1/193
13. CiteSeerX. http://citeseerx.ist.psu.edu/


