

## A Role of Semantic Web and Ontology in Information Retrieval

<sup>1</sup>Sangita S. Modi, <sup>2</sup>Dr.Sudhir B. Jagtap

<sup>1</sup>Research Scholar, Swami VivekanandMahavidyalaya, Udgir,Dt:-Latur

<sup>2</sup>Professors and Principal, S.R.T.M. University, Nanded, Maharashtra, India.S.R.T.M. University, Nanded, Maharashtra, India.

### ABSTRACT

Web Mining is an application of data mining which focuses on discovering relevant data from Web content. The Semantic Web describes a web as data rather than documents. It characterizes information in understandable manner more implicitly for humans and computers. It was developed with the help of Ontology, which is the pillar of the Semantic Web. The semantic Web depends on integration and use of semantic data, and semantic data depends on ontology. Ontology can provide a common vocabulary, a grammar for publishing data, and can supply a semantic data which can be used to preserve the Ontologies and keep them ready for inference. This also helps in personalized filtering mechanisms for users to consume relevant, interesting information from web sites. By combining web mining and semantic web, we can retrieve relevant data called as semantic web mining. This paper gives an overview of semantic web mining and their applications.

**Keywords:** semantic web, ontology, semantic web mining, personalization.

### I. INTRODUCTION

Data Mining (DM)(knowledge Discovery in databases) is the process of extraction of required pattern or information from large databases using various data mining techniques such as classification, clustering, association rule etc. which helps in various decision making. The web mining is the application of data mining through which we can extract the relevant data from web pages.[1]

The online web portals for e-learning, e-government and e-commerce became a very common part of Web. In that e-learning is one of the emerging web portal through which student can learn anywhere any time. Now a days ICT(information communication technology) plays a vital role to convert traditional education system into modern education system. Internet is one of the ICT tool which is a knowledge warehouse. Students are using different educational sites for updating knowledge, learning new technologies, enrolling different courses for their academic growth.

But while surfing on web student spending lot of time to obtain relevant web sites and data. Web mining is one of the application through which we can retrieve required educational information.

#### 1. Web Data Mining

The Web is the largest publicly accessible data source in the world. The Web has many unique characteristics, like mining useful information and knowledge, fascinating and challenging the different task. The information on the Web is noisy. The noise comes from two main sources. First, a typical Web page contains many pieces of information, e.g., the main content of the page, navigation links, advertisements, copyright notices, privacy policies, etc. For a particular application, only part of the information is useful. The rest is considered noise. To perform fine-grained Web information analysis and data mining, the noise should be removed. Second, due to the fact that the Web does not have quality control of information, i.e., one can write almost anything that one likes, a large amount of information on the Web is of low quality, incorrect, or even confusing.

Web mining is divided into three types.

1. web content mining
2. web usage mining
3. web structure mining

#### 1. Web Content Mining

It is a mining of multimedia documents, involving text, hypertext, images, audio and video information. This technique is used to extract concept hierarchies and their relations from the Web, and their automatic categorization.

#### 2. Web Structure Mining

it is used to mine documents hyperlinks, as a graph of links in a site or between sites..

#### 3. Web Usage Mining :

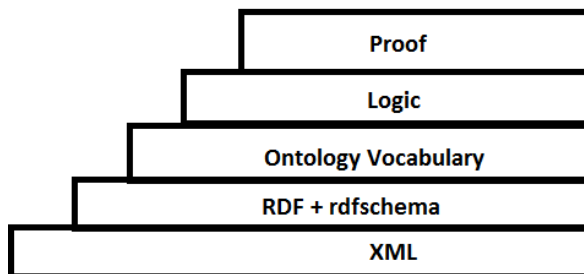
This technique mines the data generated when the user's interfaces with the Web, represented as Web server access logs, user profiles, user queries and mouse-clicks.

### 2.Semantic Web

Semantic Web is an XML application. The Semantic Web is distributed and heterogeneous service, which can be able to increase the usability of web and also ensure that its contents can be understood by machines.

Tim Berners-Lee, who invented the WWW and has worked on the Semantic Web, states that the latter "is not a separate Web but an extension of the current one, in which information is given a well-defined meaning, better enabling computers and people to work in cooperation." [2]. Thus, the Semantic Web providing a description of its contents and services in machine-readable form. It enables services to be automatically annotated, discovered, published, advertised and composed. It facilitates interoperability and the sharing of knowledge over the Web. Its main goal is to make information easily accessible and understandable by humans and computers. Figure 1 is the architecture of the Semantic Web.

- XML layer, used to characterize the structure of data.
- RDF layer, which characterizes the meaning of data.
- Ontology layer, which describes formal common agreement about meaning of data.
- Logic layer, which enables intelligent reasoning with meaningful data.



**Figure 1 Architecture of semantic web**

As semantic Web is a relatively new and dynamic field of investigation. Semantic Web having different Technologies like, XML, RDF, Metadata, ontology. The semantic Web contains resources corresponding not just to media objects like Webpages, images, audio clips, etc as the current Web does, but also objects such as people, places, organization and events. Further, the semantic Web

will contain not just a single kind of relations between resources, but many different kinds of relations amongst the different kinds of resources.

1. XML permits users to create their own tags in order to interpret Web documents. For providing and restricting the structure and content of elements contained within XML documents, XML Schema is used.
2. RDF (resource description language) is a tree structure of XML document. RDF document consists of sets of triples: a subject, predicate, and an object. For expressing data models, RDF is a language, which refers to objects and their relationships. An RDF-based model data can be described and further processed by the computers. RDF schema is a vocabulary for describing properties and classes of RDF-based resources. XML and RDF are two pillars to build an intelligent Web. RDF model uses vocabulary defined by the terms of ontology. The RDF and the XML give the computer enough information to obtain the meaning of data.[3]
3. Data about data is often called metadata. XML and RDF deal with the description of the information available on the Web.
4. Ontology was developed in AI to help in knowledge sharing and reuse, and can be built with XML and RDF. Recently many specific ontology languages have been developed. Ontologies will play a key role in supporting information exchange processes in the Semantic Web.

### **3. Ontology**

Ontology was introduced in the 1990s by AI research communities. Its aim is to integrate intelligent information, electronics commerce, knowledge management, cooperative information systems, and information retrieval. Ontology builds a backbone for artificial intelligence and semantic web purposes and they are defined by Resource Description Framework (RDF) and XML (World Wide Web Consortium (W3C)). Basically, ontology represents the relations between concepts which lie behind in a domain. The reason ontology is becoming so popular is because it promises a shared and common understanding of some domain that can be communicated between people and application systems. Because ontology aims at consensual domain knowledge, its development is often a cooperative process involving different people, possibly at different locations[4]

Common components of ontologies include:

- Objects:- objects are ground level objects or instances.
- Classes:- Sets, concepts, classes in programming.
- Relations:- relation between class and objects.
- Attributes:- set of properties, features, and parameters that object can have.
- Function terms:- Complex structures formed from certain relations that can be used in place of an individual term in a statement.
- Restrictions:- Formally stated descriptions of asserted input that must be true.
- Rules:- Statements in the form of an if-then (antecedent-consequent) sentence that describe the logical inferences that can be drawn from an assertion in a particular form.
- Axioms:- Assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application.
- Events: The changing of attributes or relations, and events.

The following steps are involved in the construction of **Ontology**:

- Acquiring the domain knowledge: Identifying and collecting appropriate information resources and present in a common language with unanimity and consistency

- Design the conceptual structure: Identify the keyconcepts of domain, theirassociated properties, and relationships among the concepts
- Develop the suitable details: Include concepts,relationships and various detail to satisfy the purpose of the ontology
- Verify: Ensure the structure for uniformity.
- Submit: Once domain expert committed by publishing it within itsplanned location.

An ontology language is a formal language used toencode the ontology. There are a number of such languages listed below.

1. DAML+OIL:-DAML stands for DARPA Agent Markup Language.DARPA in turn stands for Defence Advanced ResearchProjects Agency and is the central research and developmentorganisation for the Department of Defense.OIL stands for Ontology Inference Layer or OntologyInterchange Language. DAML+OIL are a successorlanguage to DAML and OIL that combines features ofboth. In turn, it was superseded by Web OntologyLanguage (OWL). DAML+OIL uses description logic (DL)style model theory to formalise the meaning of thelanguage.

2. SWRL :-Semantic Web Rule Language (SWRL) is a proposalfor a semantic Web rules-language, combining sublanguagesof the OWL, Web Ontology Language (OWL)DL and LITE) with those of the Rule Markup Language(Uniary/binary datalog). SWRL adds rules to OWL+DL.The reason is that these rules provide more expressivepower to description logic. SWRL plays an important rolein ontology for the semantic Web.[17]

3.Web Ontology Language:-The Web Ontology Language (OWL) is a family ofknowledge representation languages for authoringontologies. The languages are characterised by formal semantics and RDF/XML-based serialisations for thesemantic Web. The OWL was designed to add theconstructs of DL to RDF, significantly extending theexpressiveness of RDF schema both in characterizing classes and properties. It has been designed in a waythat it maps to a well-known DL with tractable reasoningalgorithms9. It has classes, sub-classes, properties, subproperties,property restrictions, and both class andproperty individuals. Standardised formal semantics andadditional vocabulary makes OWL to represent explicitterm descriptions and the relationship between entities.[18]

Ontology editors are used for creation or manipulation of ontologies there are different editors are listed below.

1.Protege[6]:-Protege is a free, open source ontology editor and aknowledge acquisition system.

Protégé is beingdeveloped at Stanford University in collaboration with theUniversity of Manchester. This application is written inJava and heavily uses Swing to create the rathercomplexuser interface

2. Knoodl:-Knoodl facilitates community-oriented developmentof OWL-based ontologies and RDF knowledge bases. It assists as a semantic technology platform, offering aJava service-based interface or a SPARQL-basedinterface so that communities can build their ownsemantic applications using their ontologies andknowledge bases.

3. DOME:-The DERI Ontology Management Environment(DOME) is developed by the Ontology ManagementWorking Group (OMWG). The job is to create a suitefor the effective management of ontologiesthat provides an integral solution for the overall problem.

4. Onto Edit:-Onto Edit is an ontology editor developed by the On-To-Knowledge project. It implements an ontologyconstructions process. The Editor stores the ontologyabstract model in the SESAME repository andproduces RDF concrete representations.

## **II. RELATED WORK ON ONTOLOGY ANDSEMANTIC WEB**

Using semantic web and ontology web information retrieval is performed for personalised web data mining. Today millions of web sites are there but to access required information, semantic web mining is useful.

P. Malo, P. Siitari, O. Ahlgren, J. Wallenius and P. Korhonen described a document filtering framework for Wikipedia using ontology, produce semantic content classifiers, using Reuters RCV1 corpus and TREC-11 filtering task definitions. Also check the performance and outperforms of content classifiers based on Support Vector Machines (SVM) and C4.5 algorithms.[7]

M. Farida Begam and G. Ganapathy proposed the theoretical framework ALMS-Adaptive Learning management System which focused on three features 1) Extracting the use's interaction, behaviour and actions and translate them into semantics, represented as Ontologies 2) Find the Learner style from the knowledge base and 3) Deriving and composing the workflow depending upon the learner style. The intelligent agents are used to perform reasoning and the personalized workflow for the e-learners.[8]

G. Deepak, J. Sheeba, and M S Hareesh Babu proposed algorithm is termed as the Differential APMI algorithm as the semantic similarity is computed twice using the Adaptive PMI strategy with heterogeneous thresholds. The PMI between a pair of terms 'm' and 'n' is given by equation (1) which is an existing PMI

$$P_{mt}(m,n) = h(m) + h(n) - h(m,n) \text{-----(1)}$$

$$APMI(m;n) = p_{mt}(m;n) / p(m)(n) + y \text{-----(2)}$$

$$Y = 1 + \log[p(m,n)] / p(n) \log[p(m)] - p(m) \log[p(n)] \text{-----(3)}$$

This method used to compute semantic similarity. The Adaptive PMI measure depicted in equation (2) is the proposed strategy for computing the semantic heterogeneity. The proposed algorithm produced an precision of 0.85, recall of 0.88. and the accuracy of 0.87 is achieved.[9]

K. Naim Shamsi, Z. Khan contributed new model for e-learning system. A metadata-based ontology proposed in model. The OWL language is used to develop ontologies. In these ontologies, the actual resources and properties specified in the RDF models are defined using the Semantic Web technology. Which contains various services and tools like registration, uploading course documents, Interactive tutorial, announcements, notifications, and simple semantic search.[10]

Su-Kyoung, Kim established inference-based web ontology, and verified inference that established web ontologies. Also can recreate new Ontology through inference. They implemented a test system to support an intelligent image retrieval based on that established web ontologies. The test result, showed with excellent performance in recall ratio and correctness rates than target system that use annotation-based ontology.[11]

Haiyang Jia, Minhong Wang, Weijia Ran, Stephen J.H. Yang e, Jian Liao a,f, Dickson K.W. Chiu they presented performance-oriented approach. Ontology is used for constructing formal and machine-understandable conceptualization of the performance-oriented learning environment. Key performance indicators (KPIs) are set up to clarify organizational training needs, and help learners create rational learning objectives.[12]

K. Sycara, Massimo Paolucci, Anupriya Ankolekar, Naveen Srinivasan proposed a vision for a Web of services which combines the growing Web services infrastructure with the Semantic Web. show that DAML-S is not just an abstract description.[13]

Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu presented a new method to offer better Web-page recommendations through semantic enhancement by three new knowledge representation models. Two new models have been proposed for representation of domain knowledge of a website. An ontology-based model which can be semi-automatically constructed, namely DomainOntoWP, and the other is

ase semantic network of Web-pages, which can be automatically constructed, namely TermNetWP. A conceptual prediction model is also proposed to integrate the Web usage and domain knowledge to form a weighted semantic network of frequently viewed terms, namely TermNavNet.[14]

The Web infrastructure currently is a distributed network of interlinked webpages with Unique Resource Locators. This helps to categorize webpages of a particular niche and identify them. The idea of Semantic Web is to push the very same infrastructure, where the linking of resources is on the data level. Semantic Web is based on the idea of Smart Data.[15]

Dbpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. Dbpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. Furthermore, it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself.”[16]

### III. CONCLUSION

This paper presents an overview of current techniques and applications of Semantic Web, Web Mining. Also given the brief explanation of the semantic web, and ontologies. Ontology tools and algorithms and their role in retrieving required data is also discussed in this paper. A framework of personalized Web mining of e-learning sites based on semantic web and ontology plays a vital role to enhance student centralized education. Personalization depending students' learning style is the key on educational field. Also a dynamic and semantic program has to be setup for minimizing the gap between traditional learning and e-learning.

### IV. REFERENCES

- [1] [https://en.wikipedia.org/wiki/Data\\_mining](https://en.wikipedia.org/wiki/Data_mining)
- [2] Berners-Lee, Tim; Hendler, James & Ora. 2002, "Lassila: The semantic web. Scient. Amer. Mag.
- [3] Decker, S. 2000. "The semantic Web roles of XML and RDF". *IEEE Internet Comp.*, 4(5), 63–74.
- [4] Gruber, T.A. 1998, "Translation approach to portable ontology specifications. Knowledge Acquisition"
- [5] Horrocks, I. 2002, "DAML+ OIL: A reasonable Web ontology language." In *Lecture Notes in Computer Science (LNCS)*, Springer-Verlag, 2287. pp. 2-13.
- [6] Knowledge modelling at the millennium the design and evolution of Protege. 1999 In *Proceedings of KAW-99*, edited by E. Grosso, H. Eriksson, R. Ferguson, S. Tu and M. Musen, Banff, Canada.
- [7] Pekka Malo, Pyry Siitari, Oskar Ahlgren, Jyrki Wallenius and Pekka Korhonen "Semantic Content Filtering with Wikipedia and Ontologies". 2010, Aalto University School of Economics Helsinki Finland
- [8] M. Farida Begam I and Gopinath Ganapathy. 2013, "Adaptive learning management system Using semantic web technologies". *International Journal on Soft Computing (IJSC)* Vol.4, No.1, February.
- [9] G. Deepak, J. Sheeba, and M S Hareesh Babu. 2016 "A Differential Semantic Algorithm for Query Relevant Web Page Recommendation". *IEEE International Conference on Advances in Computer Applications (ICACA)*. 2016 IEEE.
- [10] Khurram Naim Shamsi, Zafar Iqbal Khan, 2012, "Development of an E-Learning

- 
- System Incorporating Semantic Web". International Journal of Research in Computer Science .*
- [11] Su-Kyoung, Kim. 2007, "Implementation of Web Ontology for Semantic Web Application" *International Conference on Advanced Language Processing and Web Information Technology*, 0-7695-2930-5/07 \$25.00 © 2007 IEEE
- [12] Haiyang Jia, Minhong Wang, Weijia Ran, Stephen J.H. Yang, Jian Liao, Dickson K.W. Chiu. 2011, "Design of a performance-oriented workplace e-learning system using ontology" *Expert Systems with Applications* 38 (2011) 3372–3382 Elsevier.
- [13] Katia Sycara, Massimo Paolucci, Anupriya Ankolekar, Naveen Srinivasan. 2003, "Automated discovery, interaction and composition of Semantic Web services", Elsevier
- [14] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu. 2014, "Web-Page Recommendation Based on Web Usage and Domain Knowledge", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 26, NO. 10.
- [15] J. Hendler and D. Allemang, *Semantic Web for the Working Ontologist*, 2nd ed., Elsevier Inc., 2011.
- [16] "Dbpedia - About," *Dbpedia.org*, 2016. [Online]. Available: <http://dbpedia.org/about>.
- [17] <https://www.w3.org/Submission/SWRL>
- [18] <https://www.w3.org/OWL/>