IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 8, August 2019
ISSN (Online) 2348 – 7968
www.ijiset.com

Semantic Web and Web Mining For Ontology Building

Pallavi V Patil[1], Dr M N Nachappa[2]

[1]Assistant Professor, School of Computer Science and IT, Jain University
Bangalore, Karnataka, India
[2]Professor and Head, School of Computer Science and IT, Jain University
Bangalore, Karnataka, India

Abstract

These days huge amount of information is available on WWW. It ends in limitation of its usage for web users. It is really a tedious job to surf information or knowledge from these millions and billions of websites or webpages. There are two emerging domains that will fix this problem i.e. Web Mining and Semantic Web. These are the areas who complement each other, when used collectively, by transforming web into semantic information to make existing data sensible which is designed only for human users and afterwards users can extract knowledge from WWW. Research says that the ontology engineering is very vital to address the issue of interoperability between Web system and its knowledge which would be understandable by any user i.e. machine or human being. This paper discusses a vision of the Semantic Web for learning resources and some problems found.

Keywords— *Knowledge Management, Web Mining, Ontology, Semantic Web, Resource Description Framework.*

1. Introduction

Web plays very important role almost for everyone because it is one of the emerging and upcoming sources to reach out information and the same for future also. But the problem is with huge growth in amount of information while extracting relevant knowledge. This problem arises due to unstructured web content and lack of standardization. This issue in1 the treatment and extraction Knowledge of large volumes Web pages make hard to use information. One of the ways to fix this issue is to blend semantic web and web mining together. Techniques of data mining are used in web mining to extract relevant information from the online resources and they have been divided into three types i.e. Web content mining, Web structure mining, Web Usage Mining as shown in "Fig. 1."

Semantic Web is a recent initiative taken by Tim Berners-Lee [1] and it is an enhancement of the current Web which proposes a simpler approach to 4 'S' i.e. search, share, salvage and syndicate information. Machine-readable information is the foundation for Semantic Web and is built on capabilities of XML technology to define customized tagging schemes and Resource Description Framework's [2] flexible approach for representing data. The Semantic Web comes up with common formats for data exchange. It also records how data correlates to real time entities by permitting human user or a non-human user to commence in one database and then goes through an endless collection of knowledge databases to find different data and words to build Ontologies and representation such that it could be understood and disbursed by every users [3].

The Semantic Web is based on following criteria:

Standardized format: The Semantic Web recommends standard formats to uniform the representation of Knowledge extracted from online resources

Standardized Knowledge: It simplifies the shared knowledge in the form of ontologies by collecting information in the form of ontologies and making them available over web.

Shared Services: The usage of Web Services make possible to access applications residing on heterogenic platforms in order to resolve interoperability system.

2. WEB MINING

The web comes with set of pages that comprises of uncountable hyperlinks and huge volume of access and usage of information. Day by day amount of information on web is increasing; knowledge discovery and web mining, are becoming critical for making one's business successful in the internet world. Web mining is about discovering and analyzing relevant data from the web [4]. Data mining techniques are used by web mining for automatic discovery and extraction of information from web resources and services, it could be content, structure or usage. The most important data mining techniques applied in the web domain are Association Rule, Sequential Pattern Discovery, Clustering, Path Analysis, Classification and Outlier Discovery. Web mining can be categorized into three types, *"Fig.1,"* based on which part of the web to mine:

B. Web Content Mining:

The discovery of relevant information from the web resources (contents/data/documents) is the application of data mining techniques to content published on the Internet

[5]. The web contains many forms and types of data. Basically, the web content consists of various types of data such as plain text i.e. unstructured (image, audio, video, semi structured (Meta data as well as HTML), or structured data (XML), dynamic documents, multimedia documents.
Issues in Web content Mining are [15]:

- Hierarchical clustering.
- Predicting relationships
- Discovering grammatical rules collections.
- Finding keywords and key phases.
- Developing intelligent tools for information retrieval.
- Hypertext classification/categorization.
-

B. Web Structure Mining:

Web Structure Mining works on the web's hyperlink structure. This can offer data about page grade or trustworthiness and it also improves the search results through filtering i.e., discovers the model underlying the hyperlink structures of the web [6]. This model is proposed to study the similarities and relationships among various web sites. It also uses the web hyperlink structure as source of additional information.

C. Web Usage Mining:

Web usage mining is one of the applications of data mining techniques to discover relevant usage patterns from online resources [7], in order to understand and better aid the needs of web-based applications. This makes sense of the information generated by the web user's sessions/behaviors.

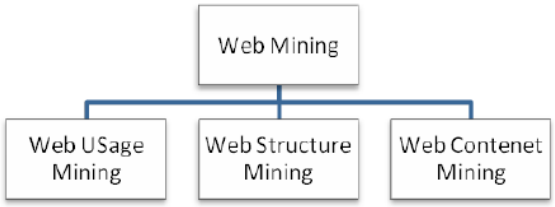

Fig . 1. : Web Mining Types

3. Semantic Web and Ontology Building

A. Semantic Web

Semantic Web is an emerging area in this era of new generation of information systems application, it enables data to be linked from one online source to any other online source and to be understood by machines so that they can achieve the assigned tasks on our behalf [1]. The main dissimilarity between Semantic Web technologies and other technologies with respect to information is that the Semantic Web is only concerned with the meaning and it is not really concerned with the structure of data. This main difference stimulates a completely different outlook on how to persist, mine, and present information. We can have different approach towards this. There are few applications which talk about huge amount of information from various sources and provides benefits as well. Remaining Applications, such as the persistence of high volumes of highly structured transactional data, do not. The Semantic Web have of four technical standards, primarily:

• ONOTOLOGY: A formal naming and definition of the types, properties, and inter-relationships of the objects that really exist for a specific area of discourse [11].

• RDF (Resource Description Framework): This gives data modeling language for the Semantic Web. All Semantic Web information storage is done and embodied in the RDF [8].

• SPARQL (SPARQL Protocol and RDF Query Language): This introduced query language of the Semantic Web. It is precisely designed to query data across several systems [9].

• OWL (Web Ontology Language): The OWL is a Semantic Web language gives rich representation and complex knowledge about things, groups of things, and relations between things.

OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer instructions, e.g. to authenticate the consistency of the knowledge or to present implicit knowledge explicit. OWL documents, known as ontologies, can be published in WWW and may address to or be addressed from other OWL ontologies. OWL is part of the World Wide Web Consortium's Semantic Web technology stack, which includes RDF, RDFS, SPARQL, etc. [10].

Though there are other standards sometimes referenced by Semantic Web literature *"Fig.2,"*, these are the foundational. The difference between a Semantic Web applications and other applications is the usage of those four technologies. However, the Semantic Web has been called many things, such as Web 3.0 or the Linked Data Web.

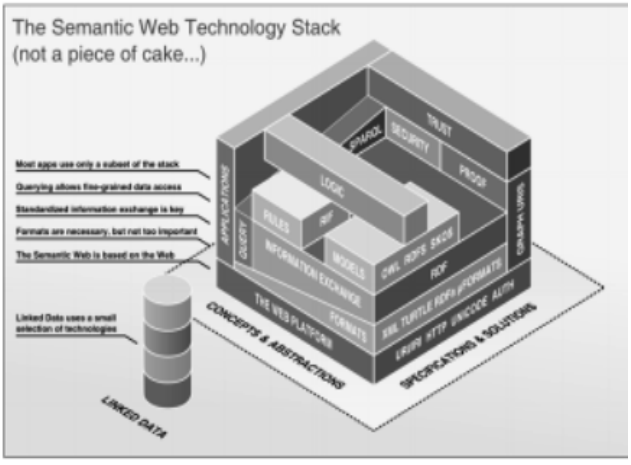

Fig .2. : Semantic Web Technology Stack

B. Ontologies Building

Ontology engineering studies the methods and methodologies for building ontologies: formal representations of a set of concepts within a domain and the relationships between those concepts. Ontology engineering is one of the regions of applied ontology [11]. Core concepts and aim of ontology engineering are also very important in conceptual modeling. As long as the area of knowledge management and knowledge sharing is concerned, ontologies have recently become famous, especially after the evolution of the Semantic Web and its supportive technologies. An ontology defines the terms and ideas used to describe and display an area of knowledge. We use many ontologies and ontology management tools.

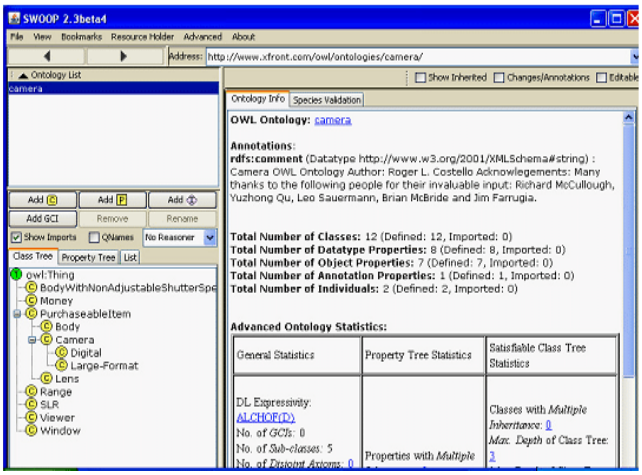

Fig .3 SWOOP with ontology example

The data which is processed using automated tools is not interpretable by software agents and can be improved by adding rich semantics to the corresponding resources [12]. One of the methodologies for the recognized conceptualization of represented knowledge domains are the usages of machine interpretable ontologies, which provide structured data in, or based on, RDF, RDFS, SPARQL and OWL. Ontology engineering is the design and formation of such ontologies, which comprises of more than just the list of terms they contain terminological, assertion, and relational axioms to define ideas (classes), individuals, and behaviors (properties). A communal way to offer the logical underpinning of ontologies is to validate the axioms with description logics, which can then be translated to any serialization of RDF, such as RDF/XML. Out there the description logic axioms, ontologies might also comprise SWRL rules [13]. This data, based on human experience and knowledge, is precious for reasons for the automated interpretation of sophisticated and ambiguous contents retrieved from web, e.g. the visual data of multimedia resources. Application areas of ontology-based reasoning consist of, but are not restricted to, information retrieval, automated scene interpretation, and knowledge discovery.

4. Semantic Web Mining

The intension behind study of Semantic Web Mining is to combine two fast developing areas of research: the Semantic Web and Web Mining. These two fields talk about the current challenges of the World Wide Web; revolving around unstructured data into a form, which machine understands, using Semantic Web tools to extract knowledge hidden in the titanic of Web data using Web Mining tools. Semantic Web Mining is a combination of these two fields, where the tools of the Semantic Web can be used to improve Web Mining and vice versa e.g. in the vast amount of data, Web Mining can find out semantic structures to build semantics for the Semantic Web. Semantic Web makes Web mining easier to achieve relevant data, but also can improve the effectiveness of Web mining [14]. The Semantic Web will make available an infrastructure that enables not just web pages, but databases, services, programs, devices, and even household appliances to both consume and produce data on the web. Semantic web mining is essentially mining the information pertaining to the semantic Web. This means mining Web pages so that the machine can better understand the information [1]. It also means mining the online sources to develop an effective semantic Web.

A. Web Pages Mining

Web pages mining is the mining, extraction and integration of useful data, information and knowledge from Web page content in order to build knowledge and to present it in form of ontology. Web content mining then play a vital role in collecting Knowledge in order to develop the semantic Web [15].

B. Ontology Mining

Web Mining in a precise domain develops and builds domain ontology [16]. For Static Web sites with static Web pages, it is realistic to develop a domain knowledge base manually or semi-manually. However, manual development and maintenance of domain ontologies necessitates a huge efforts on the part of knowledge of engineers, predominantly dynamic Website generated contents. In dynamically generated Websites, pages are usually displayed based on structured queries which are fired against databases. In such cases, we can use the database schema to directly achieve ontological information. Few of the Web servers send relevant structured data files (e.g., XML files) to web users and allows mechanisms for client-side formatting (e.g., CSS files) work out the final Web representation on client agents. In this case, it is likely to infer the schema from the structured data files [17]. When there is no direct source for acquiring domain ontologies, machine learning and text mining techniques must be employed to extract domain knowledge from the content or hyperlink structure of the Web pages *"fig 4"*. A good

representation of semantic web mining should provide machine understandability, the power of reasoning, and computation efficiency.

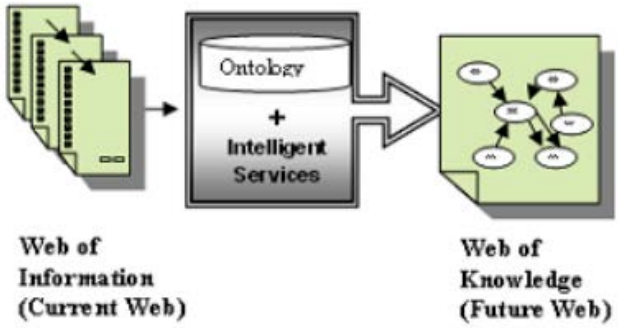

Fig .4: From Web of Information to Web of Knowledge

5. Conclusion

In this study, two fast developing research areas in World Wide Web are observed; first is web mining and second is semantic web. The combination of these two proposes new techniques to improve both areas. Web mining which is based on Semantic can improve the mining process by using the new semantic structures in the Web; and to make use of Web Mining for building up the Semantic Web. The Semantic Web also makes Web Mining much simpler because of the availability of context and Web Mining can also construct new semantic structures in the Web. The Mining assistances many areas such as e-activities, health care, bioinformatics, privacy and security, and search engines, knowledge management and information retrieval. Semantic Web Mining is an emerging research area combining Web Mining and Semantic Web. In this paper, a detailed survey of current research in Semantic Web Mining and building ontologies has been discussed. This paper also studies the integration of semantic structures in the Web to enrich the results of Web Mining and to form the Semantic Web by retaining the Web Mining techniques. We also have provided justification that the two areas Web Mining and Semantic Web requisite each other to achieve their goals especially with the role of Web Technology in commerce, education, health and government, but that the full potential of this convergence is not yet realized.

6. References

[1] T.Berners-Lee, N . Shadbolt , and W.Hall , "The Semantic Web Revisited " IEEE intelligent Systemes , pp. 96-101, 2006
[2] https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/
[3] O. Corcho , M. Fernández-López , A. Gómez-Pérez "Methodologies, tools and languages for building ontologies. Where is their meeting point" Data & Knowledge Engineering Volume 46, Issue 1, July 2003, Pages 41– 64
[4] R. Kosala and H. Blockeel "Web Mining Research A survey" SIGKDD Explorations , ACM , 2000
[5] Q. zhang, R. segall "web mining: a survey of current research, techniques, and software" international journal of information technology & decision making vol. 07, no. 04, pp. 683-720 (2008)
[6] A. abraham "Business intelligence from web usage mining" journal of information & knowledge management vol. 02, no. 04, pp. 375-390 (2003)
[7] M. aldekhail. "Application and significance of web usage mining in the 21st century: a literature review". International journal of computer theory and engineering 8:1, 41-47 2016.
[8] S. Arroyo, K. Siorpaes "ontologies and ontology languages" Handbook of Metadata, Semantics and Ontologies: 141-155.
[9] O. CURÉ , "RDF Database Systems : Chapter Three - RDF and the Semantic Web Stack " , Pages 41-80.
[10] J Pérez, M Arenas, C Gutierrez "Semantics and Complexity of SPARQL" - International semantic web conference, 2006 – Springer
[11] S Bechhofer "OWL: Web ontology language" - Encyclopedia of Database Systems, 2009 – Springer
[12] T. Bürger, M. Hausenblas "Metadata standards and ontologies for multimedia content" , Handbook of Metadata, Semantics and Ontologies: 403-439.
[13] J.Mei, H. Boley "Interpreting SWRL Rules in RDF Graphs "Original Research Article Electronic Notes in Theoretical Computer Science, Volume 151, Issue 2, 31 May 2006, Pages 53-69.
[14] Berendt, B., Hotho, A., Mladenic, D., van Someren, M., Spiliopoulou, M., Stumme,G. "A Roadmap for Web Mining: From Web to Semantic Web. Web Mining: From Web to Semantic Web" Volume 3209/2004 (2004) 1–22.
[15] Liu, B.: Web Data Mining - Exploring Hyperlinks, Contents, and Usage Data. Springer Berlin Heidelberg (2007).
[16] http://www.ontodm.com/doku.php
[17] X. Tao ; Y. Li ; N. Zhong ; R. Nayak "Ontology Mining for Personalized Web Information Gathering" published on Web Intelligence, IEEE/WIC/ACM International Conference 2007 351 – 358.
[18] Bhaskar Kapoor Savita Sharma A Comparative Study Ontology Building Tools for Semantic Web Applications International journal of Web & Semantic Technology (IJWesT) Vol.1, Num.3, July 2010.