A Study on the Issues Hindering the Semantic Web Adoption and a Proposition of an Effective Machine Learning Approach for it

Pranal Pradeep Kumar
Department of Computer Engineering
SVIT-Vasad, Gujarat, India

ABSTRACT

The web nowadays is a dynamic container of ever-increasing data. People, businesses, and devices have all become data factories that are pumping out incredible volumes of information to the web each day. There are 2.5 quintillion bytes of data created each day at our current pace on the web[1]. Also, 90% of the data produced is unstructured. Hence, there is a pressing need for the transformation of the Web 2.0 and thereby bringing in the concept of Semantic Web with a rapid pace. But, the implementation of Web 3.0 is being decelerated by various issues. However, recent advancements in the field of Machine Learning have proposed approaches to bridge the gap between semantics and the current web. In this paper, various challenges that hinder the adoption of Semantic Web and the new opportunities especially in the field of machine learning that can provide a thrust to this process have been explored.

General Terms

Semantic Web, Issues, Machine Learning Approach - Web 3.0

Keywords

Semantic Web, Ontology Mining, Machine Learning

1. INTRODUCTION

The expanding growth of information and contents across the internet has brought up a scenario that indicates that Web 2.0 is becoming inefficient. To integrate such an enormous amount of data has become a serious issue. Moreover, the format for integration that Web 2.0 uses of hyper-linking the documents is now proving to be a dull process. Rather than linking documents, it is worthwhile to interlink the data for an effective use of the World Wide Web. This is the actual concept on which the Semantic Web exists i.e. the Web of Data.

The Semantic Web can be considered as a replica of an interconnected thread of data, which requires limited human involvement in the processing of data by machines. It can be conceptualized as a lengthened variant of the present World Wide Web, and it represents a powerful way of data description in the form of a globally linked database. By advocating the incorporation of semantic content in Web pages, the Semantic Web targets the reformation of the currently available Web of unstructured documents to a Web of data.

Semantic Web also known as Web 3.0, aims at designing the web's content in a way that can be manipulated by computers meaningfully (semantically). When the word Semantic Web was coined for the first time in 2001 by Tim Berners-Lee, it was believed that in the near future, developments in it will usher in significant new functionality as machines would be able to process and "understand" the data that they merely display at present[2].

However, there are many hurdles in the present day that continue to hinder the rapid enactment of the Semantic Web. Some of the challenges are discussed in the next excerpt:

2. ISSUES CONCERNING THE SEMANTIC WEB ADOPTION

1) Data-Interlinking: A fundamental problem faced is the identification and matching of data and the capacity of determining the multiplicity of data references to the same real-world objects, by defining equivalences among data in the form of data links.

2) Amount of accessible semantic content: Contents annotated according to particular ontologies so that a concept, as well as a meaning, is provided are known as semantic contents. Since the infrastructure of the Semantic Web is still being built (RDFS, OWL, etc.), there is little semantic content available.

3) Ontology Development: The advancements in ontologies are affected mainly by three issues namely:

• Formation of kernel ontologies that all the areas use.

• Rendering methodological and technological assistance for most of the activities of the ontology development process.

• Ontology progression concerning the annotation of data.

4) Authentication and Integration of the semantic data: "Anybody can say anything on anybody" is the basic principle on which the Semantic Web carries out its working[3]. Hence to assure that a particular statement did come from an authorized source - Authentication and has not been manipulated - Integration is one of the most important issues that need to be formalized on a commercial scale.

3. THE MACHINE LEARNING PERSPECTIVE

Artificial Intelligence, Machine Learning, and whatever comes next, are phrases that are intimately associated with the very essence of the Semantic Web vision. Ironically, when the vision of Semantic Web was initially unleashed, it received a lot of criticism about being too ML and AI oriented. Many advocates the fact that instead of making data meaningful, larger focus is being made on making smarter computers and consequently the Semantic Web is losing focus.

However, Semantic Web and ML in many terms are complementary to each other[4]. And recent breakthroughs in the field of machine learning have brought many methods into light that could possibly overcome the mentioned issues of implementing Semantic Web.

Semantic Web infrastructure can be efficiently constructed by applying machine learning techniques like clustering, classification, association rules, etc. An efficient ML approach for Web 3.0 is explained further.

4. METHOD AND PROCEDURE

4.1 Ontology Management based on the dynamic approach of classification

This method aims principally at two things, namely:

1. Semi-automatic annotation of unstructured data – by MSWAF.

2. Ontology mining - by suitable ML classifiers.
The combination of the above two methods results in a series of categories that provide semantics to the available web data.

Annotated contents are the building blocks of the Semantic Web, as these markings on data and related documents support advanced searching (based on concepts), reasoning about Web resources and the information visualization based on ontology.

Here the annotation is handled by a semi-automatic procedure named MSWAF (Meteor-S Web service Annotation Framework).

This is then followed by generating a schema graph and further accurately classifying it into a knowledge graph by an effectual classifier for ontologies[5].

This approach uses the Meteor-S Web service Annotation Framework and the Naive Bayes algorithm for classifying the web contents and services based on arguments definition matching. This procedure includes the implementation of two steps which are explained further.

Step 1: Firstly, MWSAF translates the WSDL (Web Service Description Language) definitions into a common representation called a schema graph. Here the MWSAF bridges the gap between the WSDL format and ontologies by relating the bindings with concepts[6]. This step comprises pre-processing of data using textual mining techniques and then annotating the relevant extracted information.

Step 2: The second and the final step involves the Naive Bayes classifier for further categorization of the schema graph. This classifier also takes into account natural language documentation and descriptions of the web services involved in the Semantic Web. This classifier at the end deduces a sequence of candidate categories for powerful ontology mapping.

5. CONCLUSION

With the ever-expanding amount of data on the web, there is an earnest need for the implementation of the Semantic Web at a rapid pace. But, indifferent quality, limited interlinking, and limited expressiveness of mappings between related data hinder broader adoption. However, many such obstacles can be effectively tackled with a Machine Learning approach of involving an accurate annotation framework and a Machine Learning algorithm-based classifier. Also, the fact that Semantic Web and Machine Learning are complementary to each other for successful implementation of both in various domains cannot be dismissed.

6. REFERENCES

[1] Micro Focus Blog, "Data on the internet each day", https://blog.microfocus.com/

[2] Tim Berners-Lee, Mark Fischetti, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor, Harper San Francisco, 1999.

[3] W3C website, "Resource Description Framework (RDF): Concepts and Abstract Data Model", http://www.w3.org/TR/2002/WD-rdf-concepts-20020829/

[4] Devopedia,"Semantic Web." Version 3, May 29, https://devopedia.org/semantic-web

[5] Shalini Batra, Seema Bawa, "Review of Machine Learning Approaches to Semantic Web Service Discovery", Journal of Advances in Information Technology, 2010.

[6] Nicole Oldham, Christopher Thomas, Amit P. Sheth, and Kunal Verma, "METEOR-S Web service annotation framework with machine learning classification", In Semantic Web Services and Web Process Composition, volume 3387 of LNCS, pages 137–146, San Diego, CA, USA, 2004. Springer