A SEMANTIC WEB IN KNOWLEDGE-BASED DATA MINING

Mr.Ch.Srikanth Reddy[1]., Mr.K.Kranthi Kumar[2]
1,2 Assistant Professor, Department Of CSE.,
(✉srikanthcse06@gmail.Com, ✉ kurlakranthikumar@gmail.Com)
1,2 Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

Abstract

The Semantic Web provides a more advanced online service that systematically and reliably synchronizes and organizes all web-based data. Accurately picking the data needed to meet user demands and choosing them for output has proven to be a formidable challenge in the field of web-based data mining. In this study, we offer a strategy for online 3.0 data mapping. Via ontology's, and get the necessary information using a clever agent. In response to a user's inquiry, the agent returns all relevant search results for that query. Knowledge may be perceived from the information offered by the agent even when the user does not have suitable search parameters. Semantic web mining allows for the inference of such previously unknown information from pre-existing data. We provide a paradigm for web mining using intelligent agents, where the queries of users are looked for in the conventional manner (as is done, for example, by Google). The smart assistant analyzes the information that has been requested and extracts only the results that are relevant to the user's search criteria. University Faculty Information is the subject of a case study that is now in process with the purpose of evaluating the viability of the suggested model.

1. Introduction

The internet continues to develop at a breakneck pace, and it is used in a wide variety of contexts, so there is a tremendous quantity of data available online. Traditional string-based search often missed the target pages and returned several non-relevant results. A frequent complaint from users is that "Everything is on. “Internet, but we can't seem to locate it" Much of the information available online is disorganized, inconsistent, and lacking, thus [1] holds some water. The lack of connectivity between data sets increases the complexity of data mining operations. Due to the lack of defined relationships between data sets, the results of standard web mining are mostly unsatisfying in web2.0, making it almost difficult to discover previously discovered information. There has been a shift in focus on Web 3.0 as a means to better mining. Here, information is provided in a clear and organized style that facilitates collaboration between robots and humans. Semantic web data is connected to one another using ontology, allowing for efficient discovery, automation, and integration. This information can be read by computers and shared and processed by both humans and software.
The semantic web network has many layers [2] [3]. RDF [4] (Resource Description Framework) and RDF Schema offer a semantic model used to represent the Web's information and the kind thereof, and are therefore integral components of this layered architecture. SPARQL [5] is an RDF query language used to query RDF data (i.e., including statements involving RDFS and OWL [6]). Ontology vocabulary outlines the semantic connections between different types of data and

Provides a consistent definition of that data's meaning. Since it offers a machine-process able semantics and a sharable domain that may allow communication between individuals and various applications, ontology is regarded to be the backbone [7] [8] of the semantic web architecture. The idea of the Semantic Web is to improve the Web by adding data that can be understood by computers. Even while modern search engines are quite efficient, they often provide either too long or too short a list of results. The search engine's accuracy and recall may be enhanced by using machine-process able information to direct it to the most relevant sites. Extracting actionable insights and patterns from massive datasets is what data mining is all about. The goal of Web Mining is to unearth hidden patterns and associations in how information on the Web is used. As the data being mined is mostly syntactical in nature, it is difficult to find meaning using simply this data. As a result, there has been a rise in the prevalence of formalizations of the semantics of Web sites and navigational behaviour. In order to mine the information available on the Web, Semantic Web Mining integrates the two concepts of the Semantic Web. Most data on the Web are so unstructured that only humans can understand them, yet there is so much of it that only robots can analyze it effectively. Web mining attempts to solve this problem by (semi-)automatically extracting the important information concealed in this data and making it accessible in manageable amounts, while the Semantic Web works to make the data (also) machine readable. Semantic web allows knowledge mining via the web, as opposed to data mining. The capability of intelligent agent [9] helps users discover the required results for all linked keywords according to their needs. In this study, we'll examine how an agent may automatically and independently recognize all ontology items [10] relevant to a user's query request using web mining, therefore allowing the user to unearth previously undisclosed information. In Section 2 of the remainder of the work, we first provide examples of our suggested model of semantic web mining and the procedures involved in using it. In what follows, we provide a quick overview of our ongoing case study using semantic web-based representation to University Faculty Information. A smart agent may be used to learn new information with the help of ontology's, and we quickly explain how this works. Our article culminates with a brief summary of our efforts and an outline of our next steps.

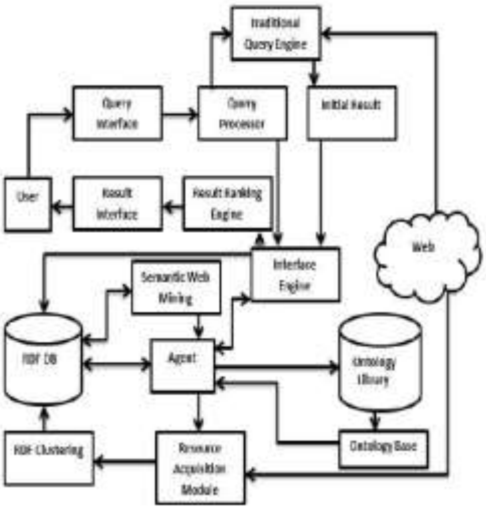

Fig. 1. Proposed web mining model under semantic agent framework

2. Proposed Model

The vast majority of data on the internet is unstructured, making it very challenging to collect and organize. By combining a structured semantic network with the unstructured real world, we take into account both the classic web mining model and the semantic web mining model enabled by a semantic agent. Condition in the network In Fig. 1 we show our suggested model, which consists of the following actions.
The first stage entails the user submitting a query request to the query processor through the query interface. In a data server, the part responsible for handling queries is called the query processor. The second step involves the query processor making concurrent calls to both conventional query engines and the intelligent agent through the interface engine with the user's request as arguments. The user may stop mining instantly with the interface stop controller. A query engine is a service that receives a query description, processes and provides the results of the query to the caller. This service interprets client-submitted search queries and conceals implementation details for the underlying data sources, serving as a go-between for the two parties. The results from the RDF database are supplied to the interface engine from the traditional query engines. The third step is to compile multiple notions about online items into a single ontology to use in agent-based searches. To collect information from the web, a specific clustering method [4] is often utilized. Expertise is combined in an ontology model, [4] which set up for early ontology construction. For later use, the ontology level will be saved in a dedicated ontology library system [4].In a fourth step, the agent queries the RDF database based on the parameters sent to it by the query processor and the interface engine. If the agent's RDF database has already cached the user's requested results, the agent will provide those results to the user immediately through the interface engine. In contrast, if the requested results are not located in the RDF database, the agent will search for all potential associations between the user's request and other web entities in the ontology library and construct an ontology foundation containing relational entities.

3. Ontology-based searching

All conceptual information about objects in the area is stored in an ontology library, which is accessible only at the ontology level. Agent will begin searching the ontology for all potential nodes connected to user supplied parameters when user contacts agent with data parameters. The completeness of the data makes this investigation feasible. On the semantic web, sets have clear definitions and are linked to one another. Agent provides the user with a wide variety of options from which to pick based on the user's specific needs. Thus, semantic web agents are preferred by users over web2.0 search engines.

Case study: University Faculty Information

Using a semantic web methodology, we created a database containing university faculty information. Forty different colleges and universities in Bangladesh were taken into account for the case study. For the sake of concision, only data pertinent to the Computer Science discipline will be evaluated. For each organization, we've encoded some basic details into Protégé. [11] The use of OWL. The faculty members' research projects, financial information, academic and industrial partnerships, publications, etc. are all encoded in a database.

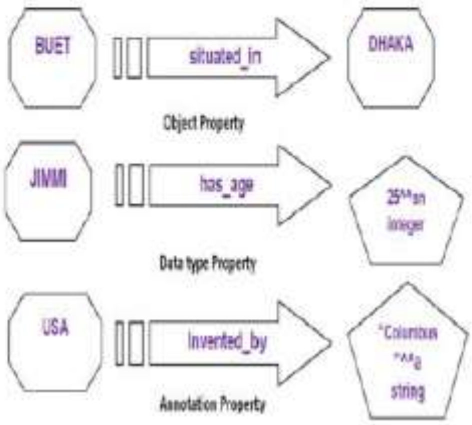

Figure 2: Many categories of OWL characteristics.

People stand in for things in the space we're exploring. In this context, "property" refers to a binary connection on persons. Set theory is used to understand OWL classes, so they may be thought of as groups of people. Formal language is used to describe them. Detailed definitions of what it takes to be a member of the class. Connections are represented in OWL by their attributes. Figure 2 depicts the three primary categories of residential real estate. Properties of objects are associations between objects, whereas properties of data types are associations between individuals and data literals. To enrich the attributes of classes, persons, and objects/data types with additional information (metadata | data about data), we may make use of annotation properties. Objects may have a wide range of different qualities, including inverse, functional, transitive, symmetrical, reflexive, etc. attributes. There are two further limitation qualities, existence and universality, that go hand in hand with these.

OWL Ontology Components

The connection between entities in ontology may be retrieved using queries. Subject-relationship-object (RDF Triple) format. It's possible to distinguish between two distinct types of user requests: Understanding Is Easily Obtained in This Way The user is looking for comprehensive details on a certain professor. All potential connections to nodes with which Edward is associated will be returned. Fig.3 depicts, for instance, all of the nodes that are connected to Edward. The Revealing of Hidden Information: For the sake of argument, let's say a user is interested in the history of Edward and John's connections. The two have no inherent connection in our encoding, however. For this case, the agent determines that Jack is the node closest to both Edward and John. Now, the agent displays every conceivable link between Edward and Jack, and Jack and John. The user may infer Edward's and John's relationship from these connections (Fig. 4).

4. Conclusion

Information overload is a major problem for modern search engines. When doing a search, the most common issue is not a lack of results but rather an overwhelming volume of results. It's not unusual for a search to return hundreds of thousands of results. Several of these results are returned because the search criteria are similar to a term on the page, even though the page itself has nothing to do with the search. Word ambiguity is a serious issue for modern search engines. Because of this inability, search engines will always return all WebPages that use this term, regardless of its intended meaning, overwhelming the user with a mix of relevant and irrelevant results.

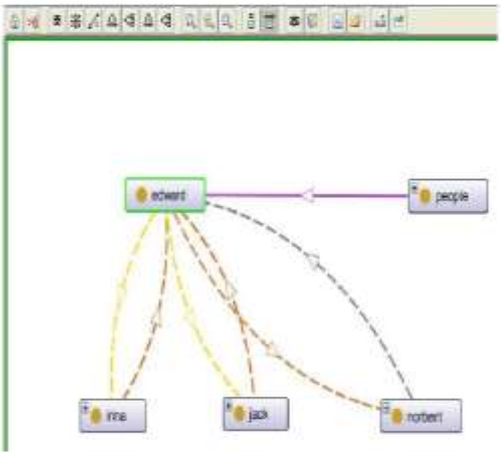

Fig. 3. Simple knowledge acquisition

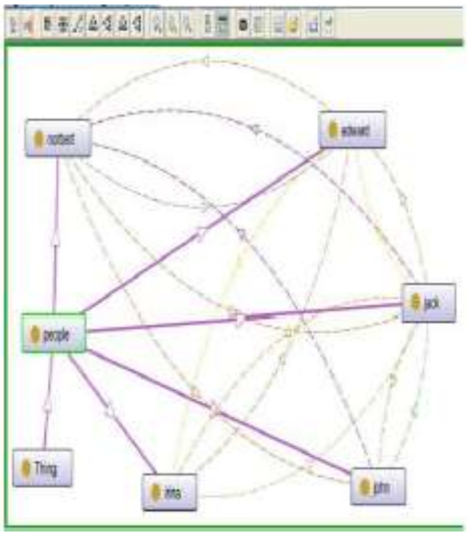

Fig. 4. Unknown knowledge discovery

With the help of the Semantic Web, we may picture a future where structured knowledge is freely and automatically transferred between agents. Agents that successfully find user-requested destination paths using ontology have access to a wide range of conveniences, including but not limited to: automation, AI, integration, machine-to-machine communication, and more. Capacity to convey ideas etc. Thanks to these tools, we can provide online users with knowledge mining rather than data mining. We've just created a model that uses a single agent and a procedure to filter away irrelevant information from a large data set, leaving just the information that was specifically requested. Coordinating different types of agents is a significant issue in any multi-agent system. Our long-term goals consist of coordinating the actions of several agents in a shared space.

Knowledge mining relies heavily on solving the problem of how to handle situations in which multi-agents acquire knowledge about their surroundings autonomously.

References

[1] M. Hepp. Semantic Web and Semantic Web Services: Father and Son or Indivisible Twins? IEEE Internet Computing, vol. 10, no. 2, pp. 85-88, March/April, 2006.

[2] W3C Semantic Web, http://www.w3.org/2001/sw/

[3] T. Berners-Lee, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by It's Inventor. 1999, Harper San Francisco.

[4] W. Yong-gui and J. Zhen. Research on semantic Web mining. 2010 International Conference on Computer Design and Applications (ICCDA), vol. 1, pp.67-70, 2010.

[5] A. Seaborne and E. Prud'hommeaux. SPARQL Query Language for RDF, W3C Recommendation, W3C, January, 2008.

[6] I. Horrocks, P. F. Patel-Schneider, F. V. Harmelen. From shiq and rdf to owl: The making of a web ontology language. Journal of Web Semantics 1 (2003) 2003.
[7] A. Grigoris and F. van Harmelen. A Semantic Web Primer. The MIT Press, Cambridge, Massachusetts April 2004.
[8] Semantic Web - XML2000, slide 10". W3C. http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10- 0.html. [Accessed on 13-05-2012]
[9] S. Russell and P. Norvig. Artificial Intelligence: A Modern Approach (3rd ed.). Prentice Hall Press, Upper Saddle River, NJ, USA. 2009
[10] A. Chakravarthy. Mining the semantic web. In: Paper Proceedings of the First AKT Doctoral Colloquium (2005).
[11] N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Fergerson and M. A. Musen. Creating semantic web Contents with protege-2000. In: Protégé-2000. IEEE Intelligent Systems, Vol. 16, No. 2, Mar 2001, pp. 60–71.