Developed an Intelligent Knowledge Representation Technique Using Semantic Web Technology

M.Samsuzzaman
Universiti Kebangsaan Malaysia, Malaysia
Email: sobuzcse@eng.ukm.my

M. Rahman[1], M.T Islam[2],T. Rahman[1], S. Kabir[1], R.I. Faruque[2]
[1]Patuakhali Science and Technology University, Bangladesh
[2]Universiti Kebangsaan Malaysia, Malaysia
Email: {hera_cse, esty_cc@yahoo.com, tariqul,rashed@ukm.my, summa.cse@gmail.com}

***Abstract*—Semantic web offers a smarter web service which synchronizes and arranges all the data over the web in a disciplined pattern. In data mining over the web, accuracy of selecting necessary data as user demand and pick them for output counts as a major key challenge from long ago. Our approach contributes a complete and automatic mapping of data over web3. 0 through ontology and accesses them by intelligent web agent. The agent offers all possible output related to user request, from which user could find desired information. When a user has insufficient data parameters to search, they can gain knowledge from the relational outputs provided by the agent and thus semantic web mining enables unknown knowledge acquisition or discovery. Here, in this paper we briefly illustrate and discuss the architecture of semantic web, then propose a model for web mining to discover knowledge under a framework of agent, and finally discusses the ways, how agent finds out user query related nodes from ontology.**

***Index Terms*—Semantic Web, Intelligent Agent, Web Mining, Ontology, Knowledge Representation.**

I. INTRODUCTION

Ongoing rapid progress and extensive application of the internet, there is a massive amount of information resources distributed on the web. The conventional string based search often misses extremely relevant pages and feedbacks a lot of irrelevant pages for user request. A common major problem for a user that "Everything is on the web, but we just cannot find what we need [1]" is partially true. Because, most of the data over the web is scattered, unstructured, often inconsistent and insufficient. Data sets are not interlinked with each other which makes mining more difficult. There are huge examples where users get bored because required information was not given on the web or they were lost. Another case is when users have very insufficient data parameters to search. To discover unknown knowledge is almost impossible in web2. 0, because there are no relationships among data sets which makes traditional web mining results almost unsatisfactory. For better performance, people are now faced toward web3. 0 which is an extension of current web2. 0. Here, information is presented on the web in a well-defined and structured manner, enable machines and human to work cooperatively. Data in the semantic web is interlinked among each other through ontology which makes effective discovery, mechanization and assimilation possible. These data has a major key factor that they are machine readable and can be shared and processed by automated tools as well as people. Intelligent agent [2] facility enables users to find desired results for all possible related terms with respect to requirements. Our work has focused on how an agent detects all possible entities from ontology during web mining [3] related to a user query request on its own in an automated manner which enables the user to discover unknown knowledge.

II. SEMANTIC WEB ARCHITECTURE

The semantic web architecture constructed by seven different levels which is organized of a layered architecture [4-5], according to Tim Berners-Lee (the inventor of semantic web). The starting layer URI and Unicode are the base for the structure of the whole system. Unicode provides a unique encoding system for processing resources. It is the universal standard encoding system for computer character representation and recognition. Before Unicode [11] computer character representation, there were several different encoding systems which made a massacre to the combination and communication across machines. Now it's so much easier.

doi:10.4304/jetwi.5.4.407-412

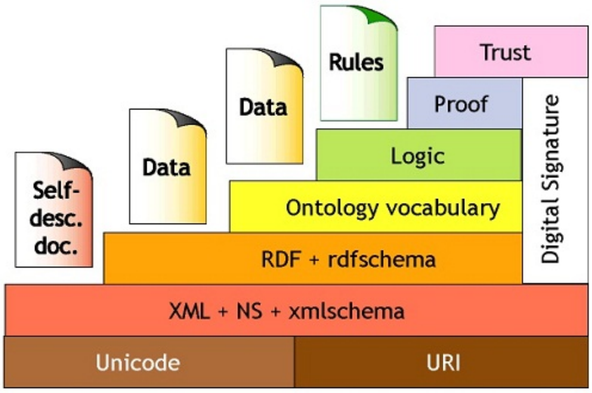

Figure 1. Semantic web architecture [4-5].

URI (Uniform Resource Identifier) provides resource documentation, which allows accurate reclamation of information possible. A resource can be anything that has an identity such as a web site,web page, a document, an image , a audio and vedio file and a person. URLs (Uniform Resource Locator), a widely used type of URIs, are very commonly used in the web, which contains the resource address. The Second layer consists of three parts- XML (Extensible Markup Language), NS (Namespace), XML Schema [8]. This layer represents linguistic data through a standard web format. XML provides a standard way for better sharing of composing information. On the other hand, it improves the freedom to structure. XML Schema pronounces the structure of XML documents. Namespaces [12] provide numerous ways to be eligible the tags and attributes in an XML document with URIs which makes them truly unique on the Web and thus, Universal (among other things). This is really important because every resource on semantic web must be identified uniquely. XML Query is a standardized language for conjoining documents, databases, web pages and almost anything else. It is very widely employed, powerful, and easy to learn. On the other hand, Namespaces [12] permits the combination of different vocabularies. For example, if a document is not marked-up and then each machine may display the documentation its own way which makes document exchange extremely challenging. XML is a mark-up language that tracks certain rules and if all documents are marked-up using XML then there is uniform representation and presentation of used documents. This is one of the significant progresses of the WWW(World Wide Web). The third layer RDF [8] (Resource Description Framework) and RDF Schema offers a semantic model used to describe the information on the Web and type. SPARQL is an RDF query language - it can be used to query any RDF-based data (i.e. including the statements involving RDFS and OWL). Querying language is essential to regain information for semantic web applications. The fourth layer Ontology [8] vocabulary layer is accountable for the definition of shared knowledge and describes the semantic relationships between the different kinds of information to disclose the semantic web between the information itself and information. Ontology is considered the pillar [12, 14] for the semantic web architecture affords a machine-processable semantics and a sharable domain which can facilitate communication between people and different applications. The rule consents proof without full logic machinery. Similar rules are those used by the production systems offered in the corresponding knowledge representation subsection. They imprisonment dynamic knowledge as a set of conditions that must be fulfilled in order to accomplish the set of consequences of the rule. The Semantic Web technology for this layer is the Semantic Web Rule Language (SWRL) The fifth layer logic [8] layer is answerable for providing axioms and inference principles to deliver the basis for intelligent services. The sixth layer proof and the seventh layer trust are liable for providing authentication [9] and trust mechanisms. In the semantic web, each object must be reformed with the corresponding change in the real world. So, for detecting any false change or attacks (may be passive or active), digital signature [9] and encryption [9] techniques are used.

Trust to consequent statements will be supported by (a) verifying that the premises come from trusted sources and by (b) relying on formal logic during deriving new information. Cryptography [10] is significant to ensure and verify that semantic web statements are approaching from a trusted source. This can be accomplished by the appropriate digital signature of RDF statements.

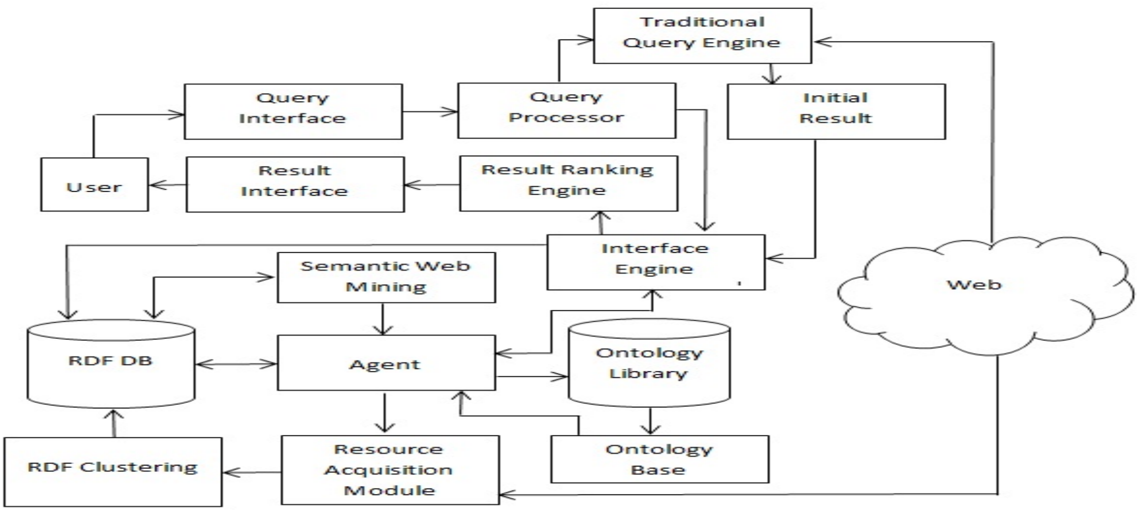

Figure 2. Web mining model under semantic agent framework

III. Agent Based Web Mining

The Semantic Web assurances to change the way of agents navigate, harvest and utilize information on the web. By providing a structured representation for articulating concepts and relationships defined by multiple ontologies, it is now possible for agents to realize users need in a better way and determine knowledge intelligently.

A. What is an Agent?

Agent [8] is an intellectual software being, which is able to complete a particular task instinctively and able for an agent to agent communications under certain state of affairs. Agents are able to perform smart analytical tasks according to the semantic information on the web which develop the accuracy of information retrieval. Rather than doing everything for a user, the agents would find possible ways, to meet user needs and offer the user choices for their achievement. Knowledge discovery from the web instead of data mining is a very intelligent process where semantic agent technology is being used.

B. Proposed Model for Web Mining by Semantic Agent

Though most of the data over the web is unstructured, it is really tough to combine or accumulate them under a common structure (construct a structured web by a night is impossible). So, we combine both traditional web mining model and semantic web mining model facilitated by a semantic agent for superior combination between well-structured semantic network and unstructured real world network situation which are shown in figure 2. The working procedure of this model is described below:

1) *The First Step:* In the beginning, the user query request is being sent to the query processor through a query interface. The query processor is the subcomponent of the data server that processes user requests.

2) *The Second Step:* The query processor calls various traditional query engines and side by side intelligent agent through the interface engine with user request parameters. Interface stop controller enables the user to shut down mining immediately if desired. A query engine [13] is a service that takes a explanation of a search request, evaluates and executes the request, and returns the results back to the caller. This service acts as an in-between layer between clients and the underlying data sources by interpreting search requests and shielding the clients from details on how to access the data sources. Traditional query engines return initial results to interface engine and results are sent to RDF database. RDF database store results in a structured way.

3) *The Third Step:* For agent based searching, an initial ontology should build and to construct this initial ontology various concepts about the objects of the web need to be gathered together. In most of the cases, specialized clustering algorithm [8] is used to gather data from the web. Ontology model merges knowledge of experts [8] in the environment to build an initial ontology. The ontology level will be warehoused in the ontology library system [8] for future levels usage.

4) *The Fourth Step:* When user request parameters are received by the agentfrom thee queryprocessorr through interface engine, agentchecks thee RDFdatabasee. If RDF database contains desired results by caching, agent directly sent results to user through interface engine. On the other hand, agent seeks out all possible relationships between user request and other web entities from ontology library and builds an ontology base with relational entities if desired results are not found in RDF database.

5) *The Fifth Step:* Ontology base contains all possible nodes related to user request collected by the agent and by acquiring knowledge from ontology base; resource acquisition module collects task related information from the web. But during the acquisition of data from the web a crucial problem arises that is arriving of irrelevant information because most of the data over the web is unstructured. The total model performance is mostly dependent on these data acquisition performance.

6) *The Sixth Step:* Resource nodes of the closest characteristics are detected and collected by resource acquisition module [8]. These nodes are being stored in the RDF database [8].

7) *The Seventh Step:* Semantic web mining module [8] mines the data in RDF database for better output and outputs is being sent to agent.

8) *The Eighth Step:* To increase the relevance of result agent performs various filtering process over the outputs of the semantic web mining module.

9) *The Ninth Step:* In this final step, all the relational results will be sent to interface engine from RDF database by the agent. Result ranking engine used for ranking the results and after ranking results will be shown to the user by a result interface. The result is given to user exhibits all possible relational aspects from which user could get desired knowledge may be known or unknown. This process is very efficient when users don't have a sufficient amount of data parameters to find desired output from the web. Instead of data mining, semantic web enables knowledge mining (knowledge acquisition) over the web. This is the main difference between web2. 0 web mining and semantic web mining.

IV. Ontology Access by Semantic Agent

Ontology level contains all conceptual knowledge about the objects in the field and stores them into an ontology library. When a user calls agent with some data parameters, agent starts to search the ontology to find all possible nodes related to user given parameters. This inquiry becomes possible because all the datasets are interlinked with each other and well defined in the semantic web. The agent gives the user a broad range of ability to choose what exactly he/she requires. Thus, users feel more comfortable to be facilitated by semantic web agent than web2. 0 search engine.

A. A Simple Ontology Building

We can construct any kind of ontology by using *Protégé* [7]. Here, we build a simple ontology named relation. Owl shown in figure 3 which has a class named people and under people some subclasses named Norbert, Jack, John, Irina and Edward.

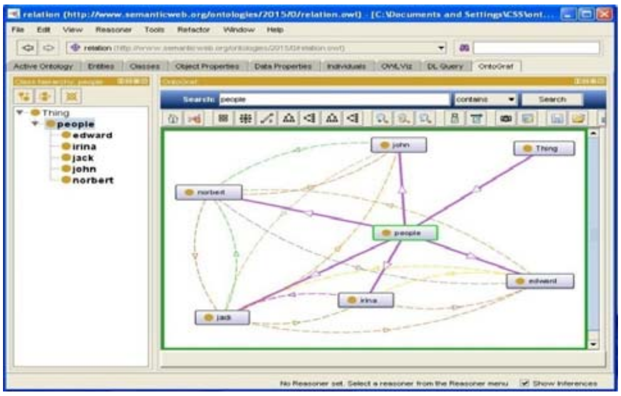

Figure 3. Relationships between entities shown in relation.Owl.

The relationship between these five people is constructed by using various object properties.

B. Agent Query

Ontology relation. The owl can be accessed by two simple queries-

// *Query for* ***some*** *property restrictions*
$query1= 'SELECT ?x ?y ?z ?a WHERE
(?x, rdfs:subClassOf, ?y),
(?y, owl:onProperty, ?z),
(?y, owl:someValuesFrom, ?a)';
// *Query for* ***all*** *property restrictions*
$query1= 'SELECT ?x ?y ?z ?a WHERE
(?x, rdfs:subClassOf, ?y),
(?y, owl:onProperty, ?z),
(?y, owl:allValuesFrom, ?a)';

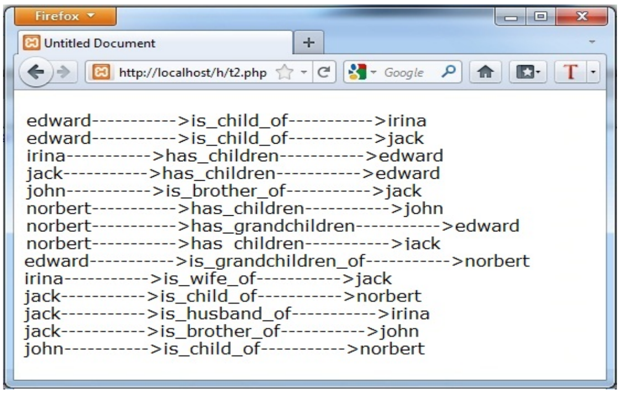

Figure 4. Relationships among all entities accessed by an agent from relation.Owl.

By executing these queries we can figure out all relationships among five people as shown in Fig 4. These relationships have always remained in RDF Triple format (Subject+ Relationship+ Object). Now, two types of user request could be found-

1) *Simple Knowledge Acquisition:* User requires total information about Edward. Relationships with all possible nodes related to Edward could be detected by SPARQL filter query shown in Fig. 6.

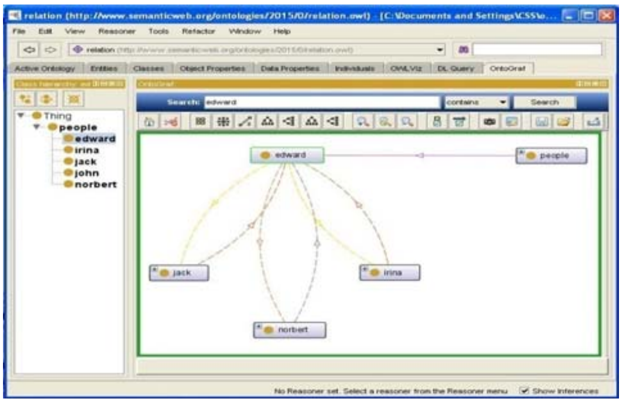

Figure 5. Relationships of Edward with all possible entities in relation.Owl.

Figure 6. Relationships of all possible nodes with Edward accessed by an agent from relation.Owl.

2) *Unknown Knowledge Discovery:* Now, time for the complex query. Let us assume that a user wants to

know about the relationships between Edward and John. From Fig. 5 it is clear that there are no direct relationship between Edward and John. In these complex situations, agent find out the closest node related to both Edward and John. From Fig. 7 we can see that the closest node is Jack. Now, agent exhibits all possible relationships between Edward-Jack and Jack-John. From these relationships user would able find out the relation between Edward and John. Ash, Edward is the son of Jack and Jack is the brother of John according to Fig. 8, then the user could obviously find out that Edward is the nephew of John and inversely John is an uncle of Edward.

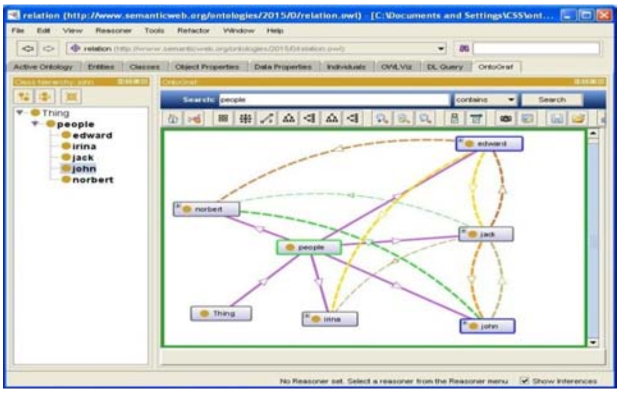

Figure 7. Relationships between Edward and John through jack in relation.Owl.

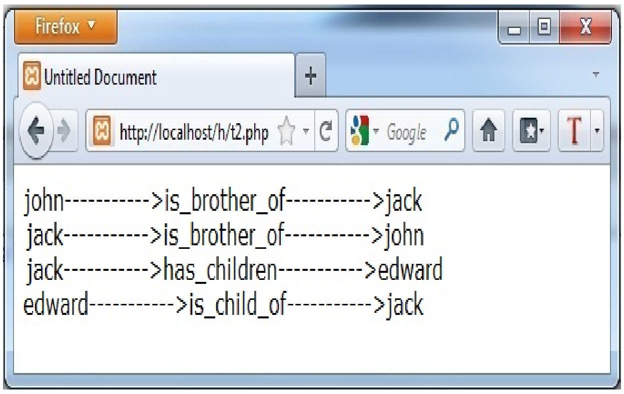

Figure 8. Agent exhibits relationships to enable user knowledge discovery.

V. Conclusion

Successfully destination path discovery by the agent through ontology as per user request provides various facilities such as automation, artificial intelligence integration, machine to machine communication ability etc. By using these facilities, we offer web users knowledge mining instead of data mining. Due to various aspects of limitations, more complex real world dealings as agent to agent communication, synchronization among multiple agents, learning from the related information in environment by agents on its own etc. which is not discussed in this paper which will be the future work.

References

[1]Martin Hepp, “*Semantic Web & Semantic Web Services*”.
[2]Stuart J. Russell and Peter Norvig, *”Artificial Intelligence: A Modern Approach“.*
[3]Ajay Chakravarthy, “*Mining the Semantic Web*”, 2004.
[4]W3C Semantic Web, http://www.w3.org/2001/sw/
[5]Berners-Lee, Tim; Fischetti, Mark (1999). *Weaving the Web*. HarperSanFrancisco. Chapter 12. ISBN 978-0-06-251587-2.
[6]Zhang Hui, ed, "*Ontology-based Semantic Web Mining Technology*", computer development and applications, 2009, 2.
[7]Matthew Horridge, *“A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools”*, Edition 1.3.
[8]WANG Yong-gui, JIA Zhen, “*Research on Semantic Web Mining*”,2010.
[9]William Stalling, “*Wireless Communications and Networks*”.
[10]www.wikipedia.com
[11]www.semanticfocus.com
[12]www.unicse.org
[13]www.tim-wellhausen.de
[14]Zhen Zhang, Bin He, Kevin Chen Chuan Chang, “*Understanding Web Query Interfaces: Best Effort Parsing with Hidden Syntax*”.

Md. Samsuzzaman was born in Jhenaidah, Bangladesh in 1982. He received the B.Sc. and M. Sc. Degree in Computer Science and Engineering from Islamic University Kushtia, Bangladesh in 2005 and 2007, respectively. Currently, he is pursuing his Ph.D. degree in telecommunication engineering in the Universiti Kebangsaan Malaysia (UKM). From February, 2008 to February, 2011, he worked as a lecturer at Patuakhali Science and Technology University (PSTU), Bangladesh. From February 2011 to till now; he is working as an Assistant Professor at the same university. He has authored or co-authored approximately 14 referred journals and conference papers. He is currently a Graduate Research Assistant at the Institute of Space Science (ANGKASA), UKM, and Malaysia. His research interests include the RF, electromagnetic field and propagation, Antenna Technology, Satellite communication, WSN and Semantic Web.

Md. Mamunur Rahman was born in Khulna, Bangladesh.He received B.Sc. in Computer Science and Engineering from Patuakhali Science & Technology University, Dhumki, Patuakhali, Bangladesh in 2012. His research interests include the Semantic web technology, Algorithm.

Mohammad Tariqul Islam was born in Dhaka, Bangladesh in 1975. He received B.Sc. and M. Sc. Degrees in Applied Physics and Electronics from the University of Dhaka, Dhaka, Bangladesh in 1998 and 2000, respectively and a Ph.D degree in telecommunication engineering from the Universiti Kebangsaan Malaysia (UKM) in 2006. In August 2000, he became an Adjunct Research Fellow at Bose Research Centre, University of Dhaka, Dhaka. He has been very promising as a researcher, with the achievement of several International Gold Medal awards, a Best Invention in Telecommunication award and a Special Award from Vietnam for his research and innovation. He has filled 6 patent applications. He has authored or co-authored 102

international journal papers and 90 international and local conference papers and 3 books. Thus far, his publications have been cited 560 times, and the H-index is 15 (Source: Scopus). He has been awarded "Best Researcher Award" in 2010 and 2011 at UKM. He served as a faculty member at the Multimedia University (MMU), Malaysia from May 2007 until May 2008. He is currently a Professor at the Institute of Space Science (ANGKASA), UKM, Malaysia. His research interests concern enabling technology for RF, antenna technology, electromagnetic absorption and radio astronomy instruments.

Tanjim Rahman was born in Barisal, Bangladesh.He received B.Sc. in Computer Science and Engineering from Patuakhali Science & Technology University, Dhumki, Patuakhali, Bangladesh in 2012. His research interests include the Semantic web technology, Algorithm, Data Mining.

Sumaiya Kabir was born in Barisal, Bangladesh.She received B.Sc. in Computer Science and Engineering from Patuakhali Science & Technology University, Dhumki, Patuakhali, Bangladesh in 2012. Her research interests include the Semantic web technology.

Mohammad Rashed Iqbal Faruque was born in Chittagong, Bangladesh in 1974. He received the B.Sc. and M. Sc. Degree in Physics from University of Chittagong, Chittagong, Bangladesh in 1998 and 1999, respectively, and a Ph.D. degree in telecommunication engineering from the Universiti Kebangsaan Malaysia (UKM) in 2012. From July 2000 to until 2007, he worked as a lecturer at Chittagong University of Engineering and Technology (CUET), Chittagong, From June 2007 to November 2008; he was an Assistant Professor at University of Information Technology and Sciences (UITS), Chittagong. He has authored or co-authored approximately 60 referred journals and conference papers. He is currently a Senior Lecturer at the Institute of Space Science (ANGKASA), UKM, Malaysia. His research interests include the RF, electromagnetic field and propagation, FDTD analysis, electromagnetic radiation, metamaterial applications and electromagnetic compatibility.