International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 02, February-2015

Customer Information Management using Semantic Web and Enhanced Data Storage Along With Tag-Based Data Retrieval in E-Commerce

Kanahiya Bodkhe, Aditi Patil , Saisha Hardikar
Department of Computer Engineering,
SKN Sinhgad Institute of Technology and Science,
Lonavala, Maharashtra, India

***Abstract* - With this paper, we try to present a new model to improve the current scenario, with utilizing concepts like semantic web and tag-based data handling techniques. The growth of conventional e-Commerce has not been at par with the high pace at which the amount of data has been growing over the internet. As a result, it fails to meet the high demands of precise data exchange activities. The proposed system differs from the conventional one as it inculcates automatic content generation based on patent search activities along with semantic annotation of information. All these renewed features transform e-Commerce into a 'Smart Space' platform to improve all the transactions happening over it.**

Keywords: Semantic Web, Tag Based Data Storage, Customer's Information Management, E-Commerce, And Intelligent Web.

1.INTRODUCTION

Let us have a closer look at how does the proposed model fulfils its purpose. Properties like ease of use, collaboration with semantic web technologies are combined by semantic wikis in the form of ontology and reasoning. They make it easy to combine such technologies which are otherwise, next to impossible. It extends a normal wiki's capability to address structured data it supports metadata in the form of semantic annotations, which correspond to an ontology that defines the properties that can be associated with different object types. The users maintains the ontology, with using knowledge models that can be represented in RDF Schema and OWL. Once created, these are available for extended queries or even adaptation of the presentation of content according to various users and domain. This also improves the process of data exchange with other applications.For example, it supports external-search functionality, such as a Web service. And the wiki system can use deductive reasoning to derive additional information for the user.

So to sum it up, semantic wikis offer:

[1] A simple formalism for semantic annotations and other kinds of content.

[2] A semantic search, by not only using keywords but also, by analyzing the relations between them.

[3] An additional automatic or semi-automatic extraction of metadata to simplify the annotation process.

Not only this, semantic wikis also cover many more applications. Features like improved navigation and search, context dependent presentation, and personalization are also added.

2. PROBLEM STATEMENT

To develop a model for Customer Information Management using semantic web and enhanced data storage along with tag-based data retrieval in e-Commerce.

This problem statement has basically 3 parts:
1. Customer Information Management
2. Sematic Web
3. Tag base data retrieval

The Customer Information Management includes the management of customers' information or about products' information which has been gathered or being delivered. It enables a complete view of a customer and their relationships across an institution, including sales and service interactions as well as the references made. It provides a safer environment on the whole as access to the private data can be restricted and it is shared with the concerned authorities through a well-defined Service Oriented Architecture (SOA).

The Semantic Web is a vision, the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications. It promises to radically improve our ability to find, sort, and classify information, tasks that consume a majority of the time spent on and off-line. The objective of the Semantic Web Architecture is to provide a knowledge representation of linked data in order to allow machine processing on a global scale. The W3C has developed a new generation of open standard markup languages that are now poised to unleash the power, flexibility and, above all, logic of the next generation of the Web, and open the door to the next generation of Web Services.

The Tag base data retrieval is used for retrieval of various kind of information stored during customer information management. The HTML provides different inbuilt tags. Using these tags the designing of the website has become so much easy. The other type of tag is Custom tags means tags which are created by the developer. These custom tags can be created in XML which provides efficient way to access to data as per developer's requirement. The use of Custom as well as predefined tags has made designing of site and retrieval of information so efficient that naïve customer as well as naive administrator can handle the site so efficiently.

3. STATEMENT OF SCOPE

A Web 3.0 search engine could find not only the keywords in your search, but also interpret the context of your request. It would return relevant results and suggest other content related to your search terms. As an example, if you typed "tropical vacation destinations under $3,000" as a search request, the Web 3.0 browser might include a list of fun activities or great restaurants related to the search results. It would treat the entire Internet as a massive database of information available for any query whereas with web 3.0, utilizing the power of semantic web will give you exact list of results, by considering the whole sentence in the query as a whole. This might increase the time taken for executing the query but instead of getting a list of millions of unrelated results, a list of query specific results will be obtained which will provide the customer with uttermost satisfaction.

Users are no doubt the first most priority when it comes to the domain of e-Commerce but customer satisfaction should not come at the cost of so much burden on the admin that it becomes impossible for him to maintain the humongous amount of data with which he has to deal with. Hence to reduce the amount of stress over the admin, this system also includes the concept of data handling techniques in which custom tags will be created for the queries which need to be executed again and again. This way, the admin will not have to rewrite the query each time the same function has to be performed. Instead, he will invoke the tag designed for that function and the work will be done in much lesser time than earlier.

Thus, the system that we propose through this paper does not only prove to be beneficial for the customer, but also for the admin.

4. PREVIOUS RELATED WORKS

Development work on exposing Web-based information to more sophisticated methods of searching and organizing has reached the point that practical applications can be created. But some of the behind-the-scenes technical work is still on going, and completing a project requires knowledge of a new array of specifications and languages. Organizations are thus left with the choice of remaining on the side lines, and potentially falling behind competitors, or entering new technological territory where there are a limited number of models of successful applications.

Today's Web is composed primarily of unstructured data, such as HTML pages. HTML pages can be searched via keyword queries, but this technology is limited. These searches cannot identify the type of information on a page. For example, they cannot determine that a string of text is a person's name or that it is the price of a product. Therefore, unlike information in a structured database, information on a Web page cannot be automatically related to information on another page, such as to extract different pieces of data about the same person and combine that information into a single personal profile. The vision of the semantic Web is to establish such capabilities.

Web 2.0 is the term given to describe a second generation of the World Wide Web that is focused on the ability for people to collaborate and share information online. Web 2.0 basically refers to the transition from static HTML Web pages to a more dynamic Web that is more organized and is based on serving Web applications to users. But the information present web 2.0 is not in a structured database, so information on a Web pages are not linked with each other, this causes extraction of unrelated information about a person or thing which is being searched it also to extract different pieces of data about the same person and combine thatinformation into a single personal profile. The vision of the semantic Web is to establish such capabilities.

5. INTRODUCTION OF THE SEMANTIC WEB TECHNOLOGY

A. Ontology

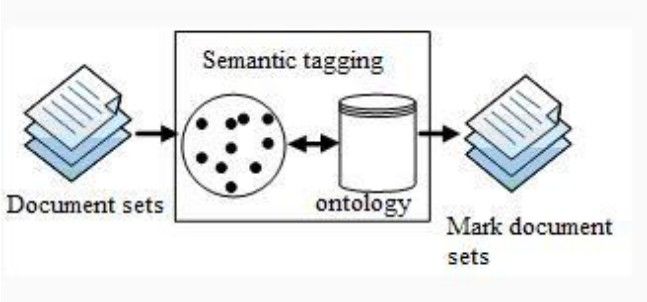
The description of the ontology about many related definitions, [3] is now recognized as one of the viewpoint, the body is refers to the sharing of the conceptual model of formal explicit instructions. The conceptual model is an abstract of the real world, so that is used to make sure these phenomenon related concepts. "Explicit" means that the concept of the type and the corresponding restriction needs a clear definition , "Formal" refers to the fact that ontology should be understood by the computer;, and "Sharing" reflecting the knowledge acquired by ontology is some common view in this field.

B. Semantic data pretreatment

Semantic data pretreatment includes the establishment of index, mark, participle and word stem completion extraction for the Chinese document. If the document is in English, the word need stem reduction, etc. The preparation workimproved retrieval efficiency and accuracy actively. The process of document semantic tagging is describing the document though Ontology concept abstracted from the document, on the other hand, the concept is the expansion and examples to the ontology actually. The user using ontology to mark the document, also is publishing process to document semantic information. It add the semantic information to the actual document. Semantic tagging process as shown in figure 1 below.

C. Semantic similarity technology

In the information processing, at present the most popular text representation method is vector space model (VSM), has been successfully applied in the famous Retrieval System SMART (System for the Manipulation and Retrieval of Text).



The basic idea of the model is using vector to represent text: (Wl, W2, W3... Wn), of which the first I Wi for a feature weights. General selection is as a feature of words. First of all text participle. A text how many words, vector have how many d. Finally the text for the vector space is said a vector, by the words as a vector of the dimension to represent text.

6.THE DESIGN OF E-COMMERCE DATABASE MANAGEMENT

A.General framework

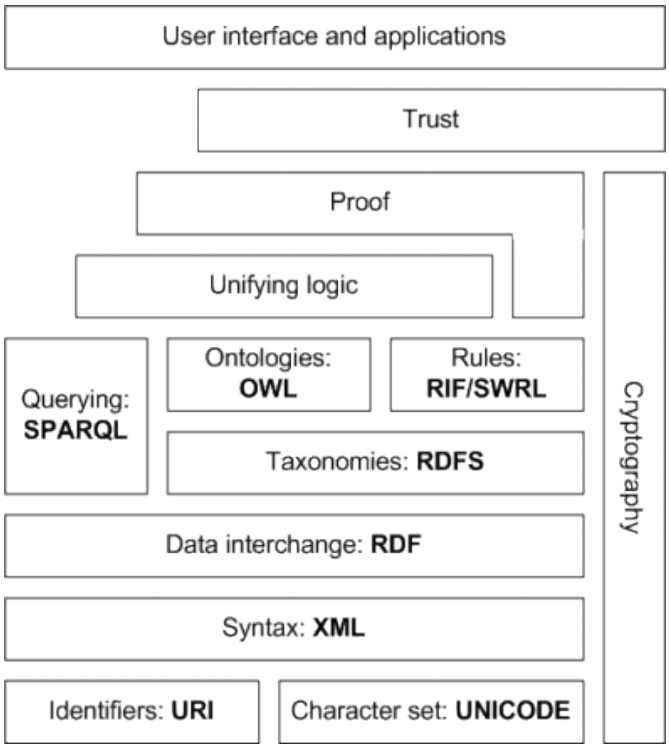


Fig1 : Proposed architecture

Extensible Markup Language (XML) layer with XML namespace and XML schema definitions makes sure that there is a common syntax used in the semantic web. XML is a general purpose markup language for documents containing structured information. A XML documents contains elements that can be nested and they may have attributes and content. XML namespace allow specifying different markup vocabularies in one XML document. XML schema serves for expressing schema of a particular set of XML documents.
A core data representation format for semantic web isResource Description Framework (RDF). RDF is aframework for representing information about resources in graph form. It was primarily intended for representing metadata about WWWresource, such as title, author, modification date of a web page, but it can be used to sorting any other data. It is based on subject-object-prediction form. All data in semantic web use RDF as a primary language. The normative syntax for serializing RDF is XML in the RDF/XML formal semantic of RDF is defined as well.

RDF itself serves as a descriptive of a graph formed by triples. Anyone can define vocabulary of terms used for more detailed description. To allow standardized description of taxonomies and other ontological constructs, a RDF Schema(RDFS) was created together with its formal semantics within RDF. RDFS can be used to describe taxonomies of classes and properties and use them to create lightweight ontologies.

More detailed ontologies can be created with Web Ontology Language (OWL). The OWL is a language derived from description logics, and offer more constructs over RDFS.It is syntactically embedded into RDF, so like RDFS, it provided additional standardized vocabulary. OWL comes in three species- OWL Life for taxonomies and simple constrains, OWL DL for full description logic support, and OWL full for maximum expressiveness and syntactic freedom of RDF. Since OWL is based on descriptive logic, it is not surprising that a formal semantics is defined for this language.

RDFS and OWL have semantics defined and this semantics can be used for reasoning within ontologies and knowledge bases described using these languages. To provide rules beyond the constructs available from these language, rule language are being standardized for the semantic web as well. Two standards are emerging – RIF and SWRL.

For querying RDF data as well as RDFS and OWL ontologies with knowledge bases, a simple protocol and RDF query language (SPARQL) are available. SPARQL is SQ-like language, but uses RDF triples are resources for both matching part of query and for returning the result of the query. Since both RDFS and OWL are built on RDF, SPARQL can be used for querying ontologies and knowledge bases directly as well. Note that SPARQL is not only query language; but also a protocol for accessing a RDF data.

7. MATHEMATICAL MODEL

Set Theory:

1. Let S= {} be as an online shopping system.

2. Identify input as Q={q1,q2,q3….qn}
   Where qi = Number of item in Q.
3. Identify input as U={U}
   Where U=User in Q.
4. Identify Das an output i.e. delivery of product.
   S= {Qn, U, D}
5. Identify Process P
   S= {Qn, U, D, P}
   P= {Pr, WIE, TID, PPP, RL}
   Where Pr=Processing of Q and Web Information
   WIE=Extracted parse information by Crawler
   TID = Identified token using token file
   PPP=Procedure programming for specific token
   RL =NLP rule for specific token
6. S={Qn, U, D, P,Pr, WIE, TID, PPP, RL }

MATHEMATICAL MODEL FOR PROPOSED SYSTEM:

1. Initialize set of Token Case
   TC= {}
2. Initialize URL
   URL= {}
3. Initialize Dictionary
   Dc= {}
4. Processes Q and Web Information contents
   Pr= {SW, ST, TN}
   Where SW= Stop word exclusion
   ST= Halting
   TN= Tokenization
5. Reading TF= {qw, sw,aw,Rl}
   Where qw= Token
   sw= Supportive symbolic arguments
   aw= Associated response threads
   Rl= NLP regulation for symbols
6. Identify token or query in queue using Dictionary Dcand Proper noun Pn
   Token = qwϵ Dc
   Proper Noun = qwϵDc/
7. Set Master vector
   Mv = {qw, sw, aw}
   Where
   qw= Query word or the Token
   sw= Supporting Query word
   aw= Related answering strings
   Mv= Master vector

Tag sentence containing Mvusing equation

$S(x)=\sum_0^n \int_0^k Mv$

n is the total number of documents

k is the total number of sentences containing master vector

S(x) is a collection of all sentences containing Mvin the entire document collection

8. Extract answer Anusing the equation

$An = \lim_{I \to K} [Tw\ (Si \epsilon Rl)]$

Where Siis a sentence in S(x)

Rlis a set of NLP rules for a specific keyword qwand

Twis the term weight

8. CONCLUSION

In this paper, we discuss in brief about Semantic Web (RDF and OWL). To serve customer effectively and efficiently, he is provided with an Intelligent Decision Making Support System, called as “Smart-Space” developed by integrating E-Commerce and Semantic Web with its Language Tools. This will improve the working of currently existing Customer Information System, as the whole process details would be handled in such a way that even naïve users will be able to handle the system with perfect ease. In future, the trend of M-commerce will also arise as customer is always in search to get appropriate product in the shortest searching time possible.

8. REFERENCES

1. S. Schaffert, IkeWiki: A Semantic Wiki for Collaborative Knowledge Management,]roc. 1st Int'l Workshop SemanticTechnologies in Collaborative Applications (Stica 06), IEEE CS Press, 2008, pp. 388-396.
2. S. Auer, S. Dietzold, and T. Riechert, _OntoWiki_a Tool for Social, Semantic Collaboration,_The Semantic Web, LNCS 4273,Springer,2007, pp. 736_ 749.
3. B. Mobasher, R. Cooley, 1. Srivastava, Automatic Personalization Based on Web Usage Mining, communications of the ACM,2000(8) , pp.l42-151
4. B. Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa, Yuqing Sun, Jim Wiltshire, Discovery of Aggregate Usage Profiles for Web Personalization , Proceedings of the Web mining for Ecommerce Workshop,Boston,2000
5. Mark Giereth and Thomas Ertl, "Visualization Enhanced Semantic Wikis for Patent Information", 12th International Conference Information Visualisation, 2008, pp. 185-190.
6. Kensaku Kawamoto, Yasuhiko Kitamura, and Yuri Tijerino, "KawaWiki: A SemanticWiki Based on RDF Templates", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06) ,2006, 0-7695-27493/ 06 IEEE.
7. G. Klyne and J.J. Carroll (ed.), "Resource Description Framework (RDF): Concepts and Abstract Syntax", W3C Recommendation, 2004
8. JSPWiki: http://www.jspwiki.org/
9. B. Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa, Yuqing Sun, Jim Wiltshire. "Discovery of Aggregate Usage Profiles for Web Personalization", Proceedings of the Web mining for Ecommerce Workshop, Boston, 2000.