

Toward semantic web using personalization techniques

Sinan Adnan Dewan

Firas Ali Hashim

Master of Science in Computer Science
Al-Mustansiriyah University

Abstract

This paper is presenting a model to make the dream of semantic web come true, it is just so far for the re-cent technology to completely move the whole internet Architecture to semantic Architecture, the model presented along this paper is to use personalization techniques to customize web servers in a way to make possessed web pages looks like it has been semantically composed. This model is a step toward a possible and meaningful synthesis of search engines and agent technology to participate in personalize customer needs and build semantic relations within the page content.

1.Introduction

The Internet is an essential medium for communicating and interacting with people worldwide. The need to publish and share information has fueled the rapid growth of the web. People is using web browser to access the web's rich content, whether for business or personal use, millions of people use web browsers to access the tremendous amount of information available on the web and to share or exchange this content with other users. Since the first appearance of Mosaic browser in 1993, the web exploded in popularity. It continued to experience tremendous growth throughout the 1990s- a period referred to as the "dot-com bubble"; that bubble burst in 2001, the users of the internet exceeded millions and the web wave hit every where and included almost every topic on the plant. [1,3]In 2003 there was a noticeable shift in how people and businesses were using the web and developing web-based applications. [3] The term Web 2.0 is the second generation of web to describe this shift of people trends. Web 2.0 became a major media buzzword and Web 2.0 companies use the web a platform to create collaborative, community-based sites. Web 2.0 involves the user -not only is the content often created by users, but users help organize it, share it, remix it, critique it, update it and so on. Web 2.0 helps users identify new meaning in already existing content through the use of tagging where there is a possibility to label web content by subject or keyword n a way that helps anyone locate information more effectively, this could be seen through different angle which is the personalization where the user can create a mapping between himself and his favorites and interests, this could be presented as the profile of

each user. This paper is introducing a personalization schema to personalize web pages through the participating of users. It is just impossible to reconstruct all these billions of web pages around the world but Web 2.0 offers create tools to deploy users effort to make this happen. [2,3,5]

The term "Semantic Web" was coined at the latest in 1998. but the issue had been discussed in 1996 and 1997, and its basic characteristics had been described in an article about SHOE (HTML based ontology description language). Semantic Web may be seen as a logical evolution of the World Wide Web. The idea is based on- the fact that nowadays, there is far too much content online on the Web for humans to find relevant information without the help of intelligent machines. The best description ever that describes Web 3.0 (Semantic Web) is " the network of meaning " while the recent Web is "network of information". [3]

2.Personalization overview

Personalization is a new term used with websites electronic-businesses which are presented on the internet. Personalization can be defined in different ways, among those definitions are: Personalization is a process of gathering and storing information about site visitors, analyzing the information, and based on the analysis, delivering the right information to each visitor at the right time. The Other definition; Personalization is information and applications matched to visitor interests and needs.[7]

Personalization has gone through different phases. Initially , personalization was used to keep the visitor on the site , exploring more of the site , which provided opportunities to advertise and promote products .The next phase attempted to increase how much money a visitor spent at each visit -by offering more expensive or related products. Today, personalization is increasingly used as a means to expedite the delivery of information to a visitor, marking the site useful and attractive to return to.[3,7]

A customer is the prime focus of any business whether it is virtual or real life. To keep the customer coming back to a website, the business has to attract the customer by providing them with what they like and prefer, personalization process makes this possible storing each, and every user's preferences and when they return they can view the website according to their needs.[7]

3.The Semantic Web (network of meaning in network of documents)

Semantic Web is the next future of the web where all the contents on the Web will be connected together based on meanings. Web developers now a days tries to draw paths that connect contents over the net such as putting voting counter on the sequence of web pages being surfed. As an example if there is a source S and it visits site A,D,G than M then there is a potential that all these sites have relations to each other, for sure , one user dose not give any

impression of that relation but when that sequence happened as frequent as come over a certain threshold then it has to be accepted as a fact that this sequence has some relation among its nodes. The advocates for the development toward the Semantic Web led by Tim Berners-Lee identify three important supporting pillars

First: semantic mark-up information suppliers

Second: intelligent software agents that capable of drawing inferences from the content

Third: computational ontologies that standardizes the communication between the agent and semantic web pages. [3,4]

4. Extensible Markup Language (XML)

The Extensible Markup Language (XML) was developed in 1996 by the World Wide Web Consortium (W3C) XML working group. XML is a widely supported open technology for describing data that has become the standard format for data exchanged between applications over the Internet. XML-based markup languages -called XML vocabularies- provides a means for describing particular types of data in standardized, structured ways. Some XML vocabularies include XHTML (Extensible Hyper Text Markup Language), MathML (for mathematics), VoiceXML (for speech), CML(chemical Markup Language-for chemistry), XBRL (Extensible Business Reporting Language-for financial data exchange), other fields' vocabularies are on the way to get finished. Massive amounts of data are currently stored on the Internet in many formats (e.g., databases, web pages, text files). Much of this data, especially which is passed between systems, will soon take the form of XML. Organizations see XML as the future of data encoding. Information technology groups are planning to integrate XML into their systems. XML has unique feature in allowing programmers to assign meaning to what would otherwise be random pieces of data. As a result, programs can "understand" the data they manipulate. [5,6]

5. The Resource description framework (RDF)

The *Resource Description Framework (RDF)* has been designed to connect Web-based data sources to the emerging Semantic Web at relatively low cost (Decker et al. 2000). According to its inventors, the RDF data model is similar to the object-oriented model, but we think that it is actually close to the functional data model (Shipman 1981). RDF distinguishes between entities, which are represented by unique identifiers (UIDs), and statements, which are valid between entities. This means that a statement connects a subject (source entity) and an object (destination entity) via a predicate/property. Figure (1) shows a schematic view of the RDF statement "The OntoAgents project has a home page at (<http://www-db.stanford.edu/OntoAgents>)". As can be seen, RDF

can be thought of as using a distinction between subject, predicate, and object.[3,6]

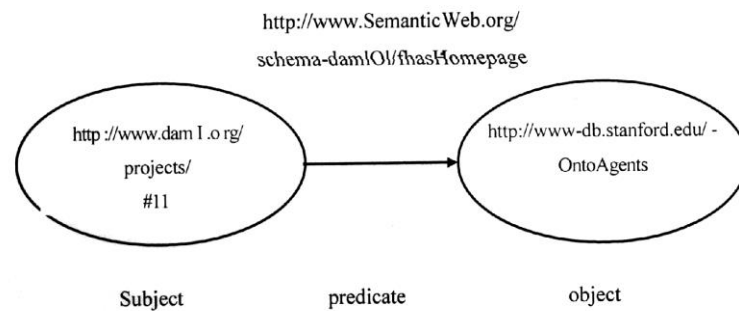


Figure (1) shows RDF graph with Subject, predicate and object

6.The Proposal system

The idea of this paper is to create XML-based descriptors, one for each web page home being retrieved from the repository, at server side and grant users the ability to markup current web page contents upon his/her experience. The following figure shows the basic idea:

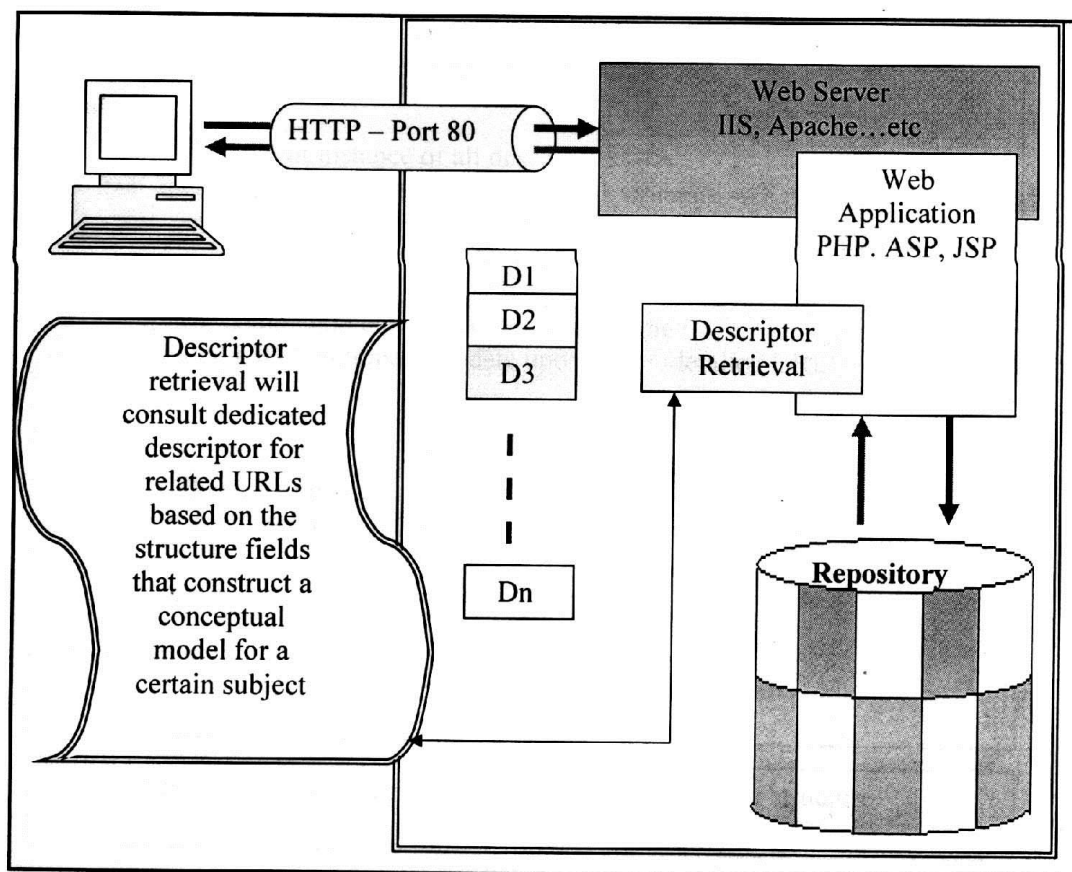


Figure (3); shows the Architecture of the proposed system

6.1. Descriptor Structure

The kernel element of the proposal suggested by this paper is the XML-based on descriptor stored at the server side. This XML-based descriptor will be referenced at every search query sent by a search engine (e.g., google, yahoo or others), and upon its field the web server can virtually construct connections based on the meaning.

```
<?xml version = "1.0"?>
<! ---- this is the root instance of all descriptors --->
<!--all descriptor will be instantiated from this structure --->
<!-- and expand upon users participations --->
<root name>
  <lup:leaf name> add child node </lup:leaf name>
  <detailed URL> automatic update upon user </detailed URL>
  .
  .
  .
  <lown:leafname> add sibling node </lown:leafname>
  < detailed URL> automatic update upon user </detailed URL>
  .
  .
  .
</root name>
```

Figure (4); shows the root instance of XML-based descriptor structure Each "leaf name" will serve as pointer to some markup section within the web page, this could be done based on the HTML tags. The most important about this descriptor is its ability to build a tree based on the "root name" and forward pointer <lup:leafname> and <lown:leafname>.

7. Conclusions

From the present paper we concludes

- 1- Semantic Web is still in its infancy stage and moving entirely to a semantic web is still far than implemented.
- 2- As this paper ended, the candidate methodology to transfer search engine mechanism toward semantic web is by constructing XML carriers and virtually modulate non-semantic constructed web pages to obtain virtual semantic content.
- 3- Web servers should have a conceptual model to percept search engine queries to develop or upgrade existing knowledge base, which later on can

be used as an intelligent reference model to reply these queries. An intelligent agent will serve best in this situation with suitable ontology.

- 4- XML is a great markup language that could be used to represent knowledge to the system. Its ancestor o:XML is fully object oriented language and can serve as a very powerful programming tool to automate web programming.

References

- 1- Andrew S. Tanenbaum, Vrije University, " Modern Operating Systems", Prentic- Hall, Inc. 2001 .
- 2- Brin, S. and L. Page, " the anatomy of a large Scale Hypertextual Web Search Engine", John Wiley & Sons Ltd. 2006 .
- 3- Gerti Kappel, Birgit Proll, siegrfried Reich, "Web Engineering", John Wiley & Sons Ltd. 2006 .
- 4- P. J. Deitel and H. M. Deitel, "Internet & World Wide Web , How to Program", Pearson Education, Inc. 2009 .
- 5- P.J. Deitel and H.M. Deitel, " Java, How to Program", Pearson Education, Inc. 2005.
- 6- P.J. Deitel and H.M. Deitel, " Java Web Services for Experienced Programmers", Pearson Education, Inc. 2003.
- 7- Senan adnan, "Design and Implementation of Personalization Website", M.S. Thesis, science college, Al - Mustansiriya university, Inc. 2006.

الخلاصة :

هذا البحث يقدم موديل لجعل حلم الويب ذات المعنى حقيقة واقعية ، ان المسافة طويلة جدا لبناء تقنية تحول معمارية شبكة الانترنت الحالية والمواقع الالكترونية فيها الى معمارية المواقع الالكترونية ذات المعنى ، ان هذا البحث يستخدم تقنيات الخصوصية لجعل الخوادم في شبكة الانترنت تحول صفحات الويب وتظهر كتكوين ذات معنى.

ان هذا البحث خطوة متقدمة باتجاه بناء محركات بحث توفر احتياجات الزبون الخاصة وبناء علاقات ذات معنى بين الزبون ومحتويات صفحة الويب .