Volume 2, No. 5, May 2011

Journal of Global Research in Computer Science

ISSN-2229-371X

REVIEW ARTICLE

Available Online at www.jgrcs.info

RESEARCH SUPPORT SYSTEMS AS AN EFFECTIVE WEB BASED INFORMATION SYSTEM

Sheilini Jindal[1*] and Gaurav Kumar[2]

[1]Department of Computer Science & Engineering

Chitkara University

Rajpura, Punjab, India

er.sheilinijindal@gmail.com

[2]Department of Computer Applications

Chitkara University

Rajpura, Punjab, India

E-mail : kumargaurav.in@gmail.com

Abstract: The aim of research support systems (RSS) is to enhance , develop and support research, Which is a major and important part of decision support systems (DSS) for scientists and researchers. Scientists are helped by Web based RSS(WRSS) in their research processes which are carried on the Web platform. WRSS are based on the assembling, integration, adaption and continuing advancement of existing computer technology and information systems for the purpose of research in the field of computers and technology. A framework of WRSS is presented by focusing on research activities and its various phases, as well as the technology support needed. The emphasis is on the conceptual formulation of WRSS. Different systems are linked to various research activities, and a mass of support sub-systems is established.The results of WRSS may lead to new and viable research tools.

Keywords: Data mining, Decision support system, Framework, Information retrieval support system, Research support system.

INTRODUCTION TO WEB BASED RESEARCH SUPPORT SYSTEMS

The World Wide Web provides a new medium for storing, presenting, gathering, sharing, processing, and using information. The impacts of the Web can be felt in most aspects of our life. The impacts are two fold: Web technology provides us with more opportunities in terms of information availability, accessibility, and flexibility. However, more challenges are in front of us. We have to find the right information and tools from largely available resources. We have to learn to use the existing tools that keep changing all the time. The study of WSS aims to take the opportunities of the Web, to meet the challenges of the Web, and to extend the human physical limitations of information processing. We define WSS as a multidisciplinary research field that focuses on supporting human activities in specific domains based on computer science, information technology, and Web technology. One of the goals is to find out how applications and adaptations of existing methodologies on Web platforms benefit our decision making and other various activities. The following are some potential benefits of Web technology:

- The Web provides a distributed infrastructure for information processing.
- The Web delivers timely, secure information and tools with a user friendly interface.
- The Web has no time or geographic restrictions. Users can access systems at any time and any place.
- Users can control and retrieve results remotely and instantly.

A TWO DIMENSIONAL VIEW OF WSS

Application domain	Technology	
	Computer technology	Web technology
Decision making	DSS	WDSS
Business application	BSS	WBSS
Information retrieval	IRSS	WIRSS
Scientific research	RSS	WRSS
Teaching	TSS	WTSS
Knowledge management	KMSS	WKMSS
Data mining	DMSS	WDMSS

Table 1: A Two Dimensional View Of Wss(Source:An Introduction To Web Based Support Systems,J.T.Yao, Department Of Computer Science, University Of Regina Regina, Saskatchewan, Canada S4s 0a2)

THE ARCHITECTURE OF WEB-BASED SUPPORT SYSTEMS

Interface, functionality, and databases are some of the components that are needed to be considered when we design a system. The architecture of WSS can be viewed as a (thin) client/server structure. The users, including decision makers and information seekers, are clients on the top layer. Internet. The interface that is designed on the server side will be presented on the client's side by browsers. The lower layers and components encapsulated by the oval dotted line are very similar to conventional computerized support systems. In other words, a Web-based support system can be viewed as a support system with the They access the system with browsers via the Web and Web and Internet as the interface.The architecture is presented from a usage point of view and is logical but not physical

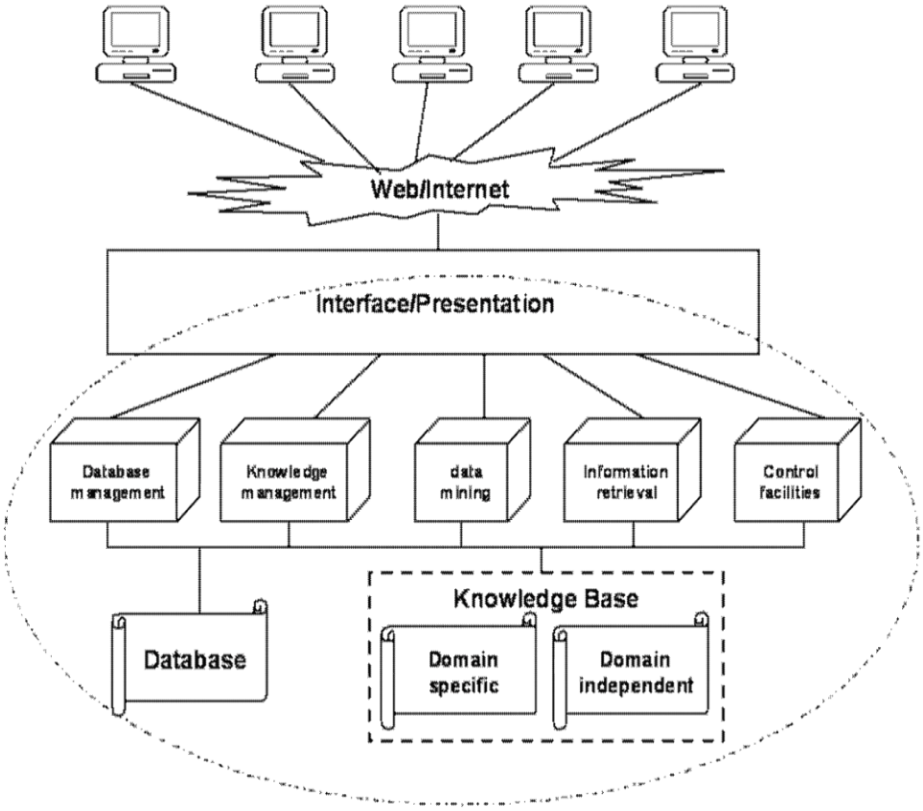

Figure 1:An architecture of web based research support system(source: http://citeseerx.ist. psu.edu)

In practice, data and control components may not necessarily sit physically on the same point of the network, which is one of the major differences between WSS and traditional computerized support systems. System components may be spread all over the network. Users of the systems are located globally. Agent, grid computing, and Web services play important roles in WSS implementation. The data layer comprises two components. A database is a basic component in any modern system. WSS is not an exception. Another major component is the knowledge base. The knowledge base stores rules, principles, and guidelines used in supporting activities. We intend to divide the knowledge base into two parts: a domain-specific knowledge base and a domain independent knowledge base. The former is the knowledge specific to the domain that is supported. The latter involves general knowledge for all support systems. Knowledge management, data management, information retrieval, data mining, and other control facilities form the management layer. These serve as middleware for the three- tier client/server architecture and as the intermediaries between the interface and data layers. Reasoning, inference, and agent technologies play important roles on this layer. The separation between the management of data and user profiles results in a secure and standardized system. To take advantage of Web technology, these processes are distributed over the Internet to form a virtual server. In fact, databases and knowledge bases on the lower tier are also distributed. The WSS can be classified into three levels. The first level is support for personal activities. An example of such support is research support for individuals. Personal research activities such as search, retrieval, reading, and writing are supported. The second level is organizational support, such as research support on an institutional level.The top level is the network level. The collaborations between organizations or decision making by a group of people like in group decision support systems fall in this level. The group-decision support room may be a virtual room on the Web.

- **RESEARCH SUPPORT SYSTEMS AND WORLD WIDE WEB**

RESEARCH SUPPORT SYSTEM FRAMEWORK

In order to explore web data, we construct a research support system framework for web data mining, consisting of four phases: source identification, content selection, information retrieval and data mining. Different phases can be explained as follows:

In the first phase, proper web sites should be chosen according to research needs. This includes identifying availability, relevance and importance of web sites. Key words searching by using search engine can be used to find appropriate web sites. After finding all web sites identified by the first phase.

The second phase is to select appropriate contents on those web sites, such as documentation, newsgroups, forums, mailing lists, etc. Usually, a web site contains many web pages, including relevant and irrelevant information. This phase is important because it decides which web information should be extracted. The selection of web pages is based on research purpose and a researcher's experience. In the information retrieval phase, a crawler is designed to automatically extract information selected during the selection phase. Specific tools and techniques are employed to effectively retrieve useful knowledge/information from web sources. Additional effort may be required for dynamic content retrieval and specfic data sources such as newsgroup, forum, etc.

The final phase is to conduct data mining on extracted web data. It includes preparing data for analysis. An extracted web page may contain missing data, extraneous data, wrong format and unnecessary characters. Furthermore, some data should be processed in order to protect privacy. Advanced data mining techniques are employed here to help analyzing data.

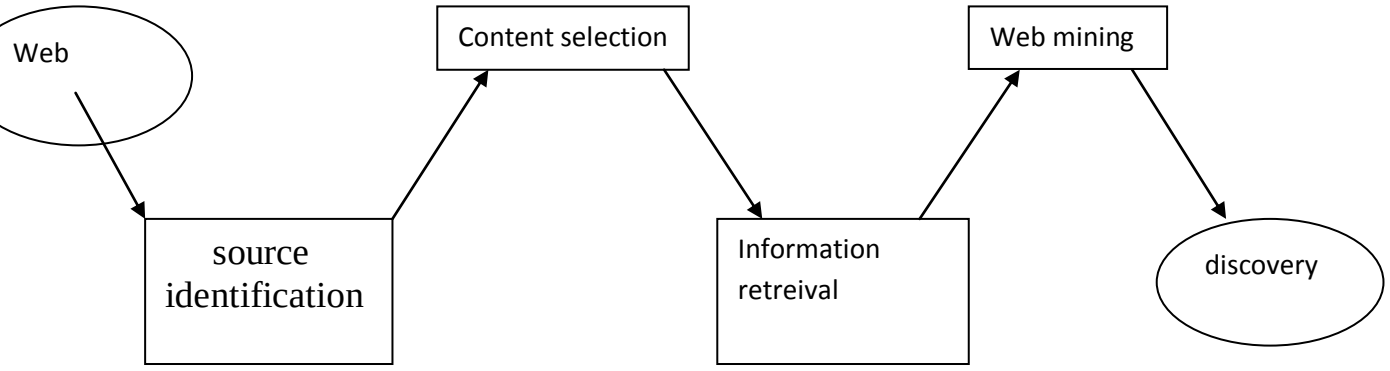

Figure 2. Research support system framework for web mining

RESEARCH FROM WORLD WIDE WEB

The advantages of the Web are often emphasized without mentioning the related difficulties. Ideally, we need to consider the problems coming with the Web, to be consistent with metaphor that the same coin has two sides. The growth of the Web, as well as information, tools, software, and services on the Web, makes scientific research easier from the point of view of easy access of information and tools. On the other hand, the limited human processing capacity becomes even more pronounced with the explosion of information, tools, and services. The opportunities and challenges offered by the Web for a scientist are summarized as follows:

• **Information on the Web:**With the fast growth of the Web and easy availability of information on the Web, we have arrived at a new information age. The Web provides a new medium for gathering, storing, processing, presenting, sharing, and using information. There is a tremendous amount of online materials, such as articles, journals, newspapers, databases, digital libraries, and so on. The easy accessibility and huge amount of information on the Web result in many difficulties for scientists, such as information overload, misinformation, fees, poorly designed navigation, retrieval, and browsing tools. How to make effective use of information on the Web becomes a serious problem. How to digest the materials on the Web and evaluate their quality are related difficult problems.

• **Web-based tools:**The advance of science and technology normally leads to new and improved tools and equipment for scientists. The development of the Web is no exception. There are also many products, tools and services on theWeb, such as news groups, downloadable software, document delivery systems, and so on. A wide spectrum and a huge number of available software systems, such as those used for data analysis, simulation, graphical representation, and document preparation are available. Those tools enable scientists to increase their research quality and productivity. With the increased number of tools, software, and services, it is a challenge to select the right tools and techniques.

It may also take more time to learn to use a new tool or software. Scientists are faced with the problem of keeping in pace with the development of new tools, which may take up their valuable research time.

DEVELOPMENT OF RESEARCH SUPPORT SYSTEMS WITH WEB MINING TECHNIQUES

“ Web mining can be broadly defined as the discovery and analysis of useful information from the World Wide Web. This broad definition on the one hand describes the automatic search and retrieval of information and resources available from millions of sites and on-line databases, i.e., Web content mining, web structure mining and web usage mining.

Association Rules

Find interesting associations or correlation relationships among a large set of data items. Basically, if X and Y are sets of items, association rule mining discovers all associations and correlations among data items where the presence of X in a transaction implies the presence of Y with a certain degree of confidence. The rule confidence is defined as the percentage of transactions containing X also containing Y .Association rule discovery techniques can be generally applied to the web mining research support system. This technique can be performed to analyze the behaviour of a given user. Each transaction is comprised of a set of URLs accessed by a user in one visit to the server. For example, using association rule discovery techniques we can find correlations in OSS study such as the following:

1. 40% of users who accessed the web page with URL/project1, also accessed/project2;
2. 30% of users who accessed /project1, downloaded software in /product1.

With massive amounts of data continuously being collected from the web, companies can use association rules to help making effective marketing strategies. In addition, association rules discovered from WWW access logs can help organizations design their web page.

Clustering

Clustering is a technique to group together a set of items having similar characteristics. Clustering is applied in the web usage mining to find two kinds of interesting clusters: usage clusters and page clusters. Usage clusters group users who exhibit similar browsing patterns. Clustering of client information or data items can facilitate the development and execution of future marketing strategies. Page clusters discover groups of pages having related content. This information is useful for Internet search engines and Web assistance providers. By using clustering, a web site can dynamically create HTML pages according to the user's query and user's information such as past needs.With clustering the goal is to separate a given group of data items (the data set) into groups called clusters such that items in the same cluster are similar to each other and dissimilar to the items in other clusters. In clustering methods no labeled examples are provided in advance for training (this is called unsupervised learning).

Classification

Classification is another extensively studied topic in data mining. Classification maps a data item into one of several predefined classes. One task of classification is to extract and select features that best describe the properties of a given class or category. In web mining, classification rules allow one to develop a profile of items belonging to a particular group according to their common attributes. For

example, classification on SourceForge access logs may lead to the discovery of relationships such as the following:

1. users from universities who visit the site tend to be interested in the page /project1;
2. 50% of users who downloaded software in /product2, were developers of Open Source Software and worked in IT companies.

CONCLUSION

This paper discusses a framework for web mining research support system and describes its procedures. It then discusses implementing techniques on web data extraction and analysis. A sourceforge web mining case is presented as an example of how to apply this framework. This work is an exploratory study of web data retrieval and data mining on web data. We try to evaluate the data extraction process and data mining software which can be used to discover knowledge in the web data. The actual interesting discoveries are still in progress. We are expected to discover interesting patterns from the data.

REFERENCES

[1] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.7261&rep=rep1&type=pdf
[2] http://www2.cs.uregina.ca/~jtyao/Conf/WSS_WIC_Summerschool.pdf
[3] http://www2.cs.uregina.ca/~wss/wss03/03/wss03-37.pdf

