Web Mining:
Strategic Web Site Design for Small Business

Mohammad A. A. ALhawamdeh
P.O.BOX (1002), Postal Code 26110,
Jarash, Jordan
962-2-6340222

Mohammad.alhawamdeh@sunderland.ac.uk
M.al-hawamdeh@lycos.com

Abstract - ***Websites are used to establish the company's policy and image, to promote and sell goods or support services and to provide customer support and to facilitate the communication and enquiries between the end users and the organization in all over the world. This research paper will describe the contribution of intelligent systems in designing the strategic website for small business to satisfy the customers need, data mining applied to the web has the potential to be quite beneficial***.

Keywords: ***Extraction, Knowledge Discovery, Web Mining.***

1. Introduction

Web mining is mining of data related to World Wide Web (WWW); the data are presented in web pages or related to web activity [8], the problems which the users face in the website are:

1- Detection of relevant information.
2- Discover of existing but 'hidden' knowledge.

When the customers log in to the website, they want to obtain information from the web [7]. So, to solve this management problem we need the right techniques and methods that derive from different areas such as: Expert Systems (ES), Artificial Intelligent (AI), DataBase (DB), and one of the Information Retrieval (IR) methods such as Structure Query Language (SQL) [11]. In brief, Web Mining: automated discovery and analysis of useful information from web documents and services using one of the data mining techniques [8].

Data mining techniques are implemented rapidly on software and hardware to maintain the information resources and integrated with new items and systems which brought online [6].

2. Benefits of Data Mining Techniques to Web Information Management

To understand the benefits of the techniques, I have to explain the business process of the company which describes the affect of the customer behaviour and the strong points of web sites to enable the firms create a strategic website.

2.1 Customer Behaviour

Customer behaviour is one of the critical success factors of any business, customer behaviour will depend on the probability of the customers coming to the company's website (visiting), this will enable the company to calculate the number of external stakeholders who are new customers, and identify the impression of the customers and their interaction with the website by calculating the number of purchases compared with the number of entered customers [2]. For example: X is small size firm for computer hardware trading, and it offers competitive prices for its products compared with others, if the number of purchases matched the expected number of operations that means the website is friendly, easy to use and based only on the required mined web data. If not, the design has some strategic problem which could not help the customer to interact effectively. Furthermore, firms can easily understand the customer requirements and behaviours by the website medium (indirect way) to get their feedbacks; that could be controlled and managed by the number of links checked, type of enquiries, Human Computer Interaction (HCI), and product information accessibility.

2.2 Website's Strong Points

Discovering the website's strong points could be in analyzing the customer feedback of the general view layout of the company's website and knowing the capability of the customers to reach all the areas of the web by the main page of the web [4] and [10]. As a small-size firm (limited resources) it would implement strategic website to match its business strategy but that should be small space website (less price) as much as possible-mining only the required web data- to save money, efforts and storage spaces (databases and secondary storage devices). Small businesses face the web data mining problem more than medium/large-size for its limited budget and resources (tangible and intangible).

3. Usage Mining to Improve the Site Usability

The reason for mining usage data within the website is to improve the site usability; the first step for the improvement process is analysis by collecting all the paths that users take, this analysis may use association rules to find the specifications between the pages as a 'sessions' which enable the company's designers to cluster the pages depend on these specifications 'similarity', this analysis will give the company opportunities to increase the probability of customer interaction [9] and [12].

3.1 Why Data Mining?

Because the complexity of the huge amount of hyperlinks and text documents, the web has multi challenges to efficient and effective sources of information and knowledge discovery:

- The complexity of text document collection, which increase the difficulty and the time of extracting data from the web and its pages (explicit knowledge), searching for the required data need a data mining technique [1] and [4].

- Internet connects a huge amount of data over several workstations reside in different locations, these data absolutely affect on decision making process of the customer when they want to buy or to make comparison with other kinds of product or services available across internet in other companies [7].

- For measuring the quality of the website, and identifying the most effective area which is include the critical data and enhance the capability of the users to retrieve the required knowledge [5].

3.2 E-Commerce Knowledge Discovery and Date Mining (KDD) Process

To illustrate the process of selling and buying via the website (online store), showing the four steps as shown in Fig.1, this research study focus on analysis and recommendations (first two steps) [7].

The analysis step will enable the developers to understand the effectiveness factors of the market and efforts of online stores, analysis will depend on action where the business analyst decisions based on customers log files and number of customers purchased, these results should be always examined by them to enable them understand the strengths and weakness areas of the web and to produce the required recommendations to keep the business on track and improve the stores bottom line [6].

3.3 Design Challenges

For designing strategic intelligent website there will be two major challenges face the designers at two levels:

1- Abstraction Level

This level include the analysis stage of gathering data about people who will affect the profit process (profit-loss account), and (market basket analysis) of the organization such as competitors, and will create traditional design of the web which enables the customers to access the immense amounts of data [3].

2- *Service Level*

At this level the organization will provide the customers with more intelligent tools (services, e.g. expert system) to communicate with website which will facilitate the extraction process and get them involved more in the system [5] and [9].

4. Web Mining Tasks

All management problems related to the design stage of the website should be solved by correct use of data mining techniques to design intelligent website which serve the business process, there are several tasks for web mining process:

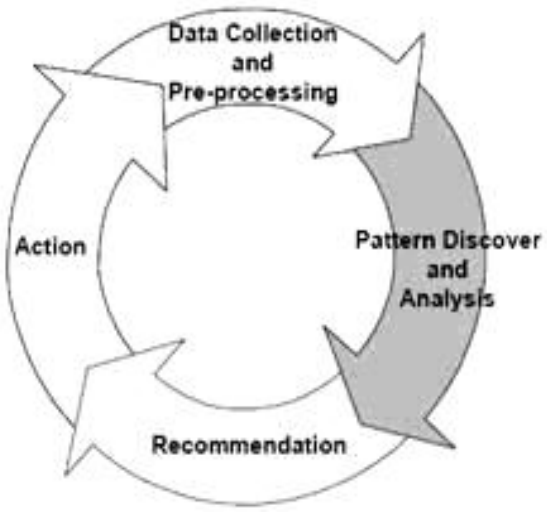

Fig.1. E-Commerce KDD Process for Web Site Design Improvement [13]

4.1 Classifying Web Documents Automatically

The website should has central search engine as intelligent tool to quickly serve the customer needs, the engine based on

keywords which classifying the documents into subdocuments to increase the speed of enquiries and save the time [5].

Typical classification model as a method for classification use unique characteristics to identify the whole training set of the business, and classify all the electronic documents by assigning the same categories of the taxonomy, there are several good typical tools could be used by developers such as Bayesian classification or keyword based association analysis [11]. Furthermore, web mining as intelligent tool could use to register every customer login to the website and directly create personal profile and classify his/her characteristics and identify his/her kind of interests [2], then classify all the customers depend on their purchases and critically enable the decision makers keep the business process well by web analysis [8] and [11].

4.2 Mining Web Page Semantic Structures and Page Contents

The designers need to know what kind of structure and semantic in a particular page could have, to enable them determine the automatic extraction of the contents, web content data consist of:

- Unstructured data: texts.
- Semi-structured data: HTML documents.
- Structured data or more: data organized in tables.

Most of the data in the business web sites are unstructured, it could be not understandable for the customers, and the application of the data mining techniques should organize and process the data and convey specific meanings of the knowledge produced to the customers. From the point of view, research in progress in web content mining can be viewed by information retrieval approaches and database [1], [9] and [11].

- Information retrieval: for improving the knowledge search of filtering for the customers, which deals with all kinds of web content data.

- Database: database related with information management applications which model the data in appropriate way on the web, to enable the customers deal with the data easy by searching based on keywords or one of the queries available on website.

The semantic web is extension to identify in which the knowledge is defined in the current website, to make the customer able to work in better cooperation [6], [7] and [12].

For example: the established website for X firm contains from five pages, where pages (P) = {a, b, c, d, e}, customers who access into the web will begin from the main page (a), and then review the rest of web pages via hypertext links to extract the wanted information. Therefore, the sessions (S) that appear would be like S = {ae, cd, bde, cecd, etc}, where (ae) an access to page number five followed directly by the access of the main page (a). There will be number of possible sessions. The sequences of those sessions will help the developers to understand both structure and linkage between pages which will help them to find out any possible opportunity to merge, expand, reduce/eliminate data, and improve the overall structure of the website. Table1 below explains the analysis process of sessions' sequences.

4.3 Mining Web Dynamics

One of the most important tasks for data mining is identifying the web dynamics which can describe the web changes and its data in its context of the semantic structure, page content, and access patterns. This task is important when the developers want to update the website, or detect the changes by storing historical information [4] and [8].

By this technique, customers can mine log records which enable them to find out more about the web pages access patterns [2], the success of this technique depend on how much and what available knowledge can we extract and get out from the raw data stored in database [11] .

Therefore developers and analysts should filter and transform the raw data into useful and information to make the extraction process easier and quicker, and they should create multidimensional view of the web depend on the URL, database, and IP address to support the customers in options interface in the web pages [1] and [5].

5. The Notion of 'Success' for Web Sites

To have successful website, the quality of the site should be designed with attentions to cover the business policy and objectives. Determining the website objectives need to measure the quality of the success to enable the analysts specifying the context, the characteristics of the objectives towards which the analysis of success factors should be accomplished appropriates to the problem specification step that outpaces any further activity in the life cycle of knowledge discovery.

Table1. Analysing Sessions to Sequences

No.	Sessions	Sequence
1	ae	(a,1) (e,1)
2	cd	(c,1) (d,1)
3	bde	(b,1) (d,1) (e,1)
4	cecd	(c,1) (e,1) (c,2) (d,1)

5.1 Extracting Knowledge (Explicit) From the Website

The research goal in content mining is extracting a useful knowledge from the web to make the process of purchasing or others easier for the customers [7], this ability requires actual understanding of the web, to enable them to create structured knowledge to facilitate the operations of retrieving the knowledge without conflict specially when the website has a huge amount of data and links [1] and [6].

One of the most effective systems for gathering the knowledge is harvest system, harvest system: a set of tools that help end users for gathering knowledge from different sources, this knowledge could be huge and has difficulties view for the customers or even for the organization's users [11].

Therefore, the developers recommending that the web site should has online analytical tools to analyze the complexity of knowledge, such as On-Line Analytical Processing (OLAP), this tool is to facilitate the understanding of complex knowledge to the users [2]and [10].

6. Conclusion

Web mining can be classified into three areas of interest based on which area of web is mined: content mining, structure mining, and usage mining. In this research paper, I surveyed the research in challenges of designing stage of strategic website beginning from general definition then moved to managerial problems which face the developers and analysts to promote and gain more customer interests, I intended to proceed the current status of web mining and its functions, with explanation to critical factors of the designing process based on external affects.

Acknowledgement

This research paper was a part of my work for Intelligent Systems for Management module at the first stage of my masters' degree of Information Technology Management (ITM) at University of Sunderland - UK, School of Computing and Technology. I would like to thank Dr. Valentina Plekhanova (ITM Programme leader) who taught me us this module and for her encouragement to participate in this event.

References

[1] B. Berendt, M. Spiliopoulou, '*Analysis of Navigation Behaviour in Web Sites Integrating Multiple Information Systems', the* VLDB Journal—the International Journal on Very Large Data Bases, Vol. 9, No. 1, pp. 56-75, Dec 1999. [Online]. Available: http://delivery.acm.org/10.1145/770000/765223/00090056.pdf?key1=765223&key2=9446029311&coll=GUIDE&dl=GUIDE&CFID=67718125&CFTOKEN=18295944
[Accessed March 2000]

[2] M. Berry, G. Linoff, *Data Mining Techniques: for Marketing, Sales, and Customer Support.* New York: John Wiley, 1997

[3] A. Berson, S. Smith, *Data Warehousing, Data Mining, and OLAP*. London: McGraw-Hill, 1997

[4] S. Chakrabarti, '*Data mining for hypertext: a tutorial survey'*, ACM SIGKDD Explorations Newsletter, Vol. 1 No.2, pp. 1-11, Jan 2000. [Online] Available: http://delivery.acm.org/10.1145/850000/846187/p1-chakrabarti.pdf?key1=846187&key2=8202839311&coll=GUIDE&dl=GUIDE&CFID=67915529&CFTOKEN=88358647
[Accessed January 2000]

[5] M. Dunham, *Data Mining: Introductory and Advanced Topics.* New Jersey: Prentice Hall/ Pearson Education, 2003.

[6] O. Folorunso, A. O. Ogunde, '*Data Mining as a Technique for Knowledge Management in Business Process Redesign*', Electronic Journal of Knowledge Management (EJKM), Vol. 2, No. 1, pp. 33-44, 2003. [Online] Available: http://www.ejkm.com/volume-2/v2i1/v2-i1-art4.htm
[Accessed September 2004]

[7] R. Kohavi, F. Provost, *Applications of Data Mining to Electronic Commerce.* Boston: Kluwer Academic Publishers, 2001.

[8] R. Kosala, H. Blockeel, '*Web Mining Research: A Survey'*, ACM SIGKDD Explorations Newsletter, Vol. 2, No. 1, pp. 1-15, Jun 2000 .[Online] Available From: http://delivery.acm.org/10.1145/370000/360406/p1-kosala.pdf?key1=360406&key2=2412986711&coll=GUIDE&dl=GUIDE&CFID=67716741&CFTOKEN=83034916
[Accessed June 2000]

[9] G. Linoff, M. Berry, *Mining the Web: Transforming Customer Data into Customer Value.* New York: John Wiley, 2001

[10] Y. Ma, B. Liu, C. K. Wong, '*Web for data mining: organizing and interpreting the discovered rules using the Web*', ACM SIGKDD Explorations Newsletter, Vol. 2, No. 1, pp. 16-23, Jun 2000. [Online] Available: http://delivery.acm.org/10.1145/370000/360408/p16-ma.pdf?key1=360408&key2=5112839311&coll=GUIDE&dl=GUIDE&CFID=67915529&CFTOKEN=88358647
[Accessed June 2000]

[11] S. Sirmakessis, *Text Mining and its Applications: Results of the NEMIS Launch Conference.* Berlin: Springer-Verlag, 2004

[12] M. Spiliopoulou, C. Pohle, *'Data Mining For Measuring and Improving the Success of Web Sites'*, Journal of Data Mining and Knowledge Discovery, Vol. 5 , No. 1-2, pp. 85-114, Jan-Apr 2001.

[13] J. Lee, M. Podlaseck, E. Schonberg, R. Hoch, '*Visualization and analysis of Click-stream data of online stores for understanding web merchandising*', Journal of data mining and knowledge discovery, Vol. 5, No. 1-2. pp. 59-84, Jan-Apr 2001.