Improving Effectiveness of Online Teaching
(An Application of Web Usage Mining)

Sartaj Ahmad
Faculty of Computer Science
Jazan University,Jazan
Kingdom of Saudi Arabia

M.Z. Khan
Faculty of Computer Science
Jazan University,Jazan
Kingdom of Saudi Arabia

ABSTRACT

WWW is one of the most valuable resources of information retrievals and knowledge discoveries and it supports Online teaching learning process. We can use this resource to upload information about particular subject (under course) and can find popularity of such information.

We can extract knowledge from this resource by many mining techniques to improve the performance of this information.

In this paper we are providing an introduction of web mining and review of web mining categories. Then we are focusing on one of these categories the web usage mining.

After this we are showing how this category can be used to raise the performance of our information.

General Term: Online teaching, Web usage mining

Keywords: WWW, Online teaching, Online learning, web mining, web usage mining

1. INTRODUCTION:

WWW has become one of the most widespread information resources. However, it demonstrates many differences to traditional information containers such as databases in schema, volume etc. Those differences making it challenging for fully use web information in an effective and efficient manner. We know huge volume of data is available on the internet. Before mining we need to gather web documents. Secondly web pages should be extracted and represented into some format as they are semi structured. Thirdly web information tends to be of diversity in meaning, training or testing data set should be large enough. Links among web pages are important resource for such type of activities. In this paper we firstly provide a survey on web mining then pay special attention on web usage mining. In the last we are using web usage mining for our teaching learning process. Finally we conclude this paper.

2. AN OVERVIEW OF WEB MINING:

Web Mining can be classified into three categories [1, 4] web content mining, web structure mining and web usage mining as shown in the following figure 1.
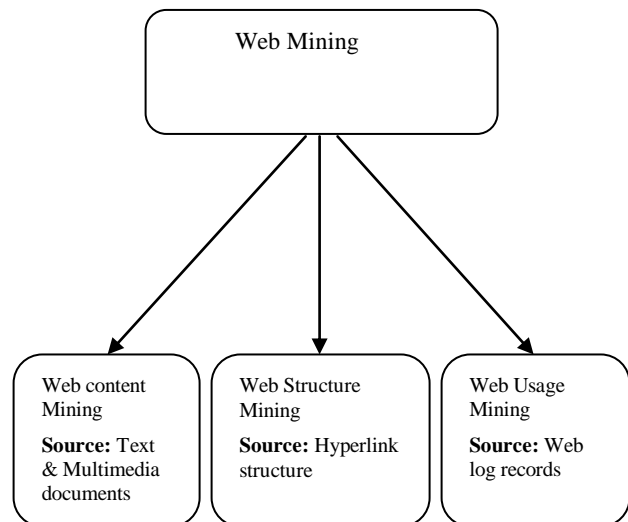

Figure 1: Web mining categories

We provide a brief introduction about each category.

2.1 Web Content Mining: This includes knowledge discovery from text documents and multimedia documents such as images, videos, audios which are embedded in or linked to the web pages [1, 6, 11].

2.2 Web Structure Mining: It deals with the structure of the hyperlinks within the web itself. This includes a newly emerging research area called Link Mining [7] .The objects in the WWW are web pages and links are in and out. In this section we review two main approaches HITS concept and PageRank method. These approaches focus on the link structure of the web to find importance of the web pages.

2.2.1 *HITS Concept*: In this concept Kleinberg [13] identifies two kinds of pages from the web hyperlink structure one authorities (pages with good source of contents) and hubs (pages with good source of links). Kindly see the following figures to understand a non negative authority weight $X[p]$ and a non negative hub weight $Y[p]$.
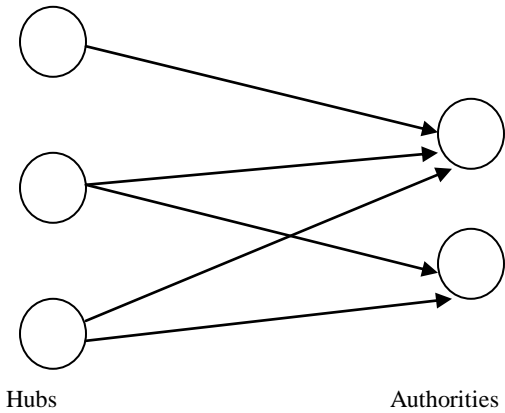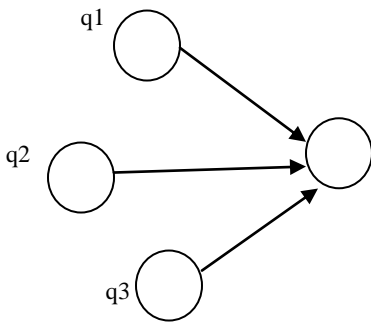
Hubs   Authorities

Figure 2 A: Set of Hubs and Authorities

$X[p]$ = sum of $y[q]$, for all q pointing to p

Figure 2 B: X- weight (from [14])

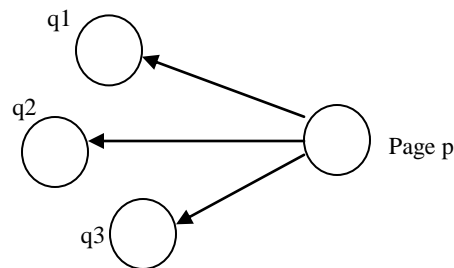$Y[p]$ =sum of $x[q]$, for all q pointing to by p

Figure 2 C: Y- weight (from [14])

2.2.2   *PageRank representation*: This algorithm is proposed by L.Page and S.Brin[9,16] and calculates the importance of web pages using link structure of the web. In this algorithm [16] we assume page A has pages P1, P2,…..Pn which point to it(i.e. are citations). There is a parameter'd' called damping factor and its value can be set between 0 and 1. We usually set d to 0.85. The number of links going out of page A is defined by C (A). The PageRank of a page is given as follows.

PageRank(A)=(1-d)+d(PageRank(P1)/C(P1)+....+ PageRank(Pn/C(Pn))

Here note PageRank form a probability distribution over web pages, so the sum of all web pages, PageRanks will be one.

2.3   Web Usage Mining:

Data for this type of mining can be collected at the different sources like server level collection, client level collection and proxy level collection. This consists of the following tasks [4] as mentioned in figure 3.
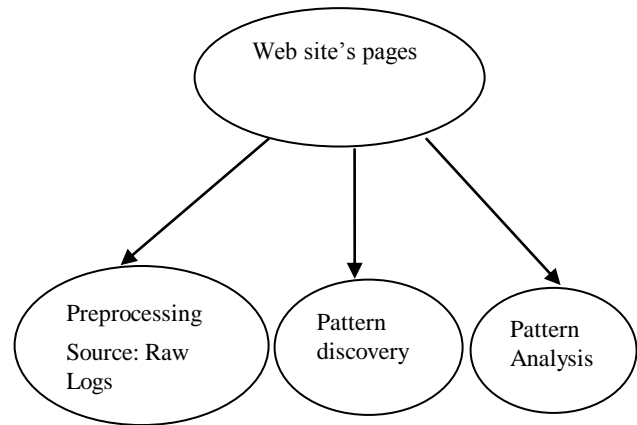

Figure 3: Web Usage Mining Process

***Content Preprocessing*:** It consists of converting the image, text, scripts and other files like multimedia into forms that are useful for such mining. Usually this comprises performing content mining such as classification or clustering.

***Pattern Discovery*:** It consists of methods and algorithm developed from several fields such as statistics, data mining, machine learning and pattern recognition. It comprises many techniques like statistical analysis, association rules, clustering, classification, sequential patterns, dependency modeling etc.

***Pattern Analysis*:** The purpose of pattern analysis is to filter out interesting rules or patterns from the set found in the pattern discovery phase.

2.3.1   *Application of Web Usage Mining*:

We use this technique for data available on Jazan University's log file. We filter this data and arrange in the form of a table as given in table 1. Then we apply statistical and time series analysis to find some interested sample to know the understanding of the students about C++ language.

Table 1:

S.No.	Topic	No. of hits/week	Session time (Hrs.)/week
1	Introduction	150	281.5
2	Looping	155	193.25
3	Array	153	286
4	Structure	120	175
5	Class	160	266
6	Pointer	80	166

Average No. of hits/week/topic = 136.33

Average session time (Hrs.)/week/topic = 227.95

Sample Analysis:

1. Pointer topic is less visited by the students because No. of hits/week is less than average no. of hits/week/topic
2. Pointer topic is less read by the student because session time (Hrs.)/week is less than Average session time (Hrs.)/week/topic.

Action: We organize a test of 10 marks for 50 students based on pointer and get the following result as in table 2.

Table 2: Test 1 Result

>=8(Greater than & Equal to 8)	>=5&<8(Between 5 & 8)	<5(Less than 5)
5	20	25

Root Cause Analysis:

As shown in the above table the no. of students having less than 5 marks is more. It means **pointer** topic is less understandable among students .It may be due to not good teaching or matter available on site.

Corrective / Prevention Action:

We need to give more emphasis on this topic. We arrange extra classes and give more exercises on this topic. We also revise matter available on this site especially for pointer.

Improvement:

We organize test2 to confirm our corrective/preventive action and get the following result as mentioned in table 3.

Table 3: Test 2 Result

>=8(Greater than & Equal to 8)	>=5&<8(Between 5 & 8)	<5(Less than 5)
10	35	5

Result comparison of test1 and test2:
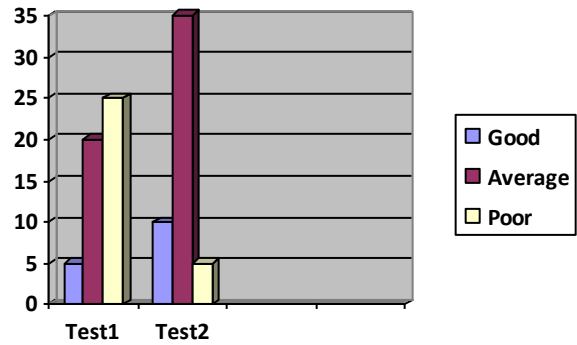
This is shown in the following figure 4.

Figure 4: Between test1 and test2 result

Finally we reduce number of poor students in pointers as shown in the above graph.

3   CONCLUSION

In this paper we survey the area of web mining. We categorized it and focused on one of its popular category web usage mining. After this we discussed its application to understand this type of mining. Since this is a large area and a lot of work to do. We hope this paper could be a useful source to improve teaching learning process. This will also help to understand web mining for identifying opportunities for further research work.

REFERENCES:

[1] Raymond Kosala, Hendrik Blockeel, Web Mining Research: A Survey, *ACM SIGKDD Explorations Newsletter*, June 2000, Volume 2 Issue 1.

[2] O. Etzioni. The world wide web: Quagmire or gold mine. *Communications of the ACM*, 39(11):65-68, 1996.

[3] Jidong Wang, Zheng Chen, Li Tao, Wei-Ying Ma, Liu Wenyin, Ranking User's Relevance to a Topic through Link Analysis on Web Logs, *WIDM' 02*, November 2002.

[4] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pag-Ning Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *ACM SIGKDD Explorations Newsletter,* January 2000, Volume 1 Issue

[5] G. Piatetsky-Shapiro, and W.J. Frawley, Knowledge Discovery in Databases. *AAAI/MIT Press*, 1991.

[6] A. A. Barfourosh, H.R. Motahary Nezhad, M. L. Anderson, D. Perlis, Information Retrieval on the World Wide Web and Active Logic: A Survey and Problem Definition, 2002.

[7] L. Getoor, Link Mining: A New Data Mining Challenge. *SIGKDDExplorations*, vol. 4, issue 2, 2003.

[8] Q. Lu, and L. Getoor. Link-based classification. In *Proceedings of ICML-03*, 2003.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd. The Pagerank citation ranking: Bring order to the web. *Technical report*, Stanford University, 1998.

[10] S. Chakrabarti, B. E. Dom, D, Gibson, J. Kleinberg, R. Kumar, P.Raghavan, S. Rajagopalan, and A. Tomkins. Mining the Link Structure of the World Wide Web. February, 1999.

[11] Wang Jicheng, Huang Yuan, Wu Gangshan, Zhang Fuyan. Web mining: knowledge discovery on the Web. *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE international Conference* - on Volume 2, Page(s):137 - 141 vol.2 - 12-15 Oct. 1999

[12] Han, J., Kamber, M. Kamber. Data mining: concepts and techniques. Morgan Kaufmann Publishers, 2000.

[13] Kleinberg, J.M., Authoritative sources in a hyperlinked environment. *In Proceedings of ACM-SIAM Symposium on Discrete Algorithms, 1998*, pages 668-677 – 1998.

[14] Cooley, R.; Mobasher, B.; Srivastava, J.; Web mining: information and pattern discovery on the World Wide Web. *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE* International Conference. Page(s):558 – 567 - 3-8 Nov. 1997.

[15] Bianchini, M.; Gori, M.; Scarselli, F.; Inside PageRank. ACM Transaction on Internet Technology (TOIT), Volume 5 Issue 1 – February, 2005.

[16] Brin, S.; Page,L. The Anatomy of a Large-scale Hypertextual Web Search Engine. *Proceedings of the Seventh International World Wide Web Conference*, 1998.

[17] http://www.google.com/. Last accessed 05/04/2010.

[18] http://www.Jazan.edu.sa Last accessed 06/04/2010.