

IMPROVED SERVER RESPONSE USING WEB PRE-FETCHING: A REVIEW

Veena Singh Bhadauriya, Bhupesh Gour, Asif Ullah Khan
Department of Computer Science & Engineering
Technocrats Institute of Technology, Bhopal, India

ABSTRACT

Web application is most common used application all over the world in order to perform communication. There are various challenges in web application like security, time and space etc. As far as time is convent web server process the request and then generates the reply to the client. In this duration time taken by web server should be minimum. This review paper throws some light on the methodologies by which the server response time can reduces. The web Pre-fetching is one of the best concepts in order to make the web application more efficient. This paper also discusses about the web pre-fetching over various methods.

KEYWORDS: *Web Mining, Web Prefetching, Web Caching*

I. INTRODUCTION

Rapid growth of web application has increased the researcher's interests in this era. All over the world has surrounded by the computer network. There is a very useful application call web application used for the communication and data transfer. An application that is accessed via a web browser over a network is called the web application. Web caching is a well-known strategy for improving the performance of Web based system by keeping Web objects that are likely to be used in the near future in location closer to user. The Web caching mechanisms are implemented at three levels: client level, proxy level and original server level [1,2]. Significantly, proxy servers play the key roles between users and web sites in lessening of the response time of user requests and saving of network bandwidth. Therefore, for achieving better response time, an efficient caching approach should be built in a proxy server.

Web caching and prefetching are the most popular techniques that play a key role in improving the Web performance by keeping web objects that are likely to be visited in the near future closer to the client. Web caching can work independently or integrated with the web prefetching. The Web caching and prefetching can complement each other since the web caching exploits the temporal locality for predicting revisiting requested objects, while the web prefetching utilizes the spatial locality for predicting next related web objects of the requested Web objects [1]. Prefetching is used as an attempt to place data close to the processor before it is required, eliminating as many cache misses as possible. Caching offers the following benefits: Latency reduction, Less Bandwidth consumption, Lessens Web Server load. Prefetching is the means to anticipate probable future requests and to fetch the most probable documents, before they are actually requested. It is the speculative retrieval of a resource into a cache in the anticipation that it can be served from the cache in the near future, thereby decreases the load time of the object [3].

Web caching is usually transparent to the user and to the application designer, except for the possible improvement in response time [4]. The application designer, when planning the development of a system, usually will not have enough information to judge if a web cache is involved. Also, if this developer is not knowledgeable in network protocols, he or she will focus on the application

functionality, i.e., the interface between the scripting language and the database and the assembly of the pre-defined response pages [5].

A pre-fetching and caching ratio model used to improve hit ratios of accessed documents, the architecture of which consists of three functional a mining mechanism consisting of the pattern mining, here in this frequent item set is found by using graph approach and based on it frequent pattern are discover, based on this caching and pre-fetching ratio is found out and Finally make conclusions.

The main Aim of this paper is to demonstrate that web pre-fetching is an effective solution to reduce web latency perceived by the users, and that it can be implemented easily and efficiently in the current real environment.

This paper has divided into six major sections including this one. First section is introduction of the topic. The second section describes the web mining with their types. Third and fourth section gives an idea about web prefetching and its types. Fifth section throws some light on previous work as a related work and finally we conclude the paper in section six.

II. WEB MINING

Web mining is the extension of data mining research [5] in the Web environment. It aims to automatically discover and extract information from Web documents and services [6]. However, Web mining is not merely a straightforward application of data mining. New problems arise in Web domain and new techniques are needed for Web mining tasks. The World-Wide Web is huge, diverse, and dynamic, and thus raises the issues of scalability, the problems of modelling multimedia data and modelling temporal Web respectively. Due to these characteristics of WWW, we are currently overwhelmed by information and facing information overload [7]. Users generally encounter the following problems when interacting with the Web [8]:

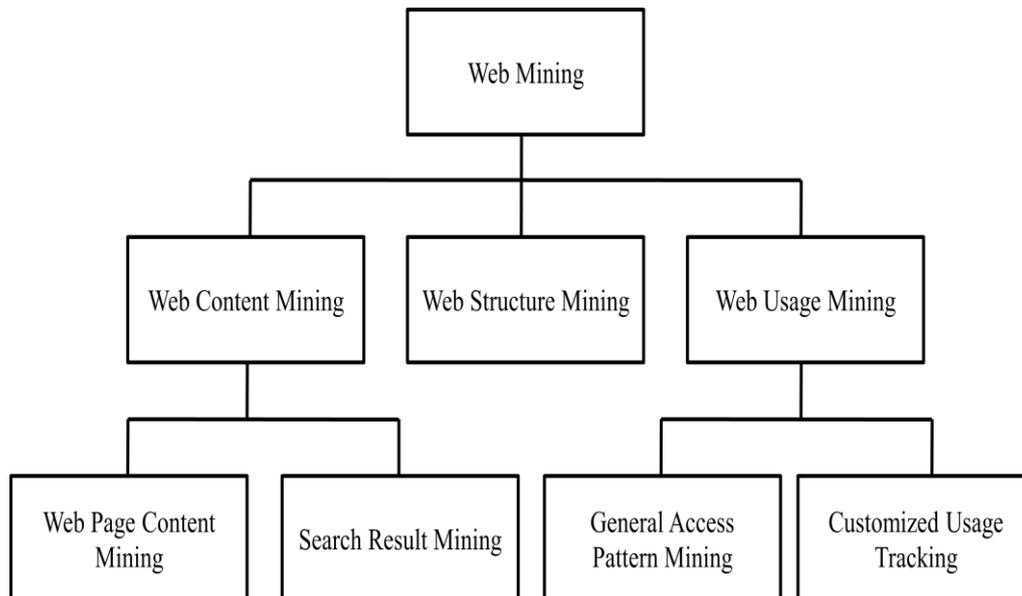


Figure 1: Taxonomy of Web Mining

- Finding relevant information:** Users can either browse the Web manually or use automatic search service provided by search engines to find the required information in WWW [9]. Using the search service is much more effective and efficient than manual browsing. Web search service is usually based on keyword query and the query result is a list of pages ranked by their similarity to the query. However, today's search tools have the problems of low precision and low recall [5]. The low precision problem is due to the irrelevance of search results and it results in the difficulty of finding relevant information, while the low recall problem is due to the inability to index all the

available information on Web, and it results in the difficulty of finding the unindexed information that is relevant.

- **Creating new knowledge out of the information available on the Web:** - Based on the collection of Web data on hand, users always wonder what they can extract from it. That is, users hope to extract potentially useful knowledge from the Web and form knowledge bases. Recent research [10] focused on utilizing the Web as a knowledge base for decision-making.
- **Personalization of the information:**-Users prefer different contents and presentations while interacting with the Web [6]. In order to attract more Web users, Web service providers are motivated to provide friendlier interface and more useful information according to users tastes and preferences.
- **Learning about consumers or individual users:**-Some Web service providers, especially the e-commerce providers, have kept a large number of records of their customer's behavior when they visit the Web sites [5,10]. Analyzing these records allow them to know more about their customers, and even predict their behaviour. To meet this need, some traditional data mining techniques are still useable, while some new techniques are created.

References [8] categorize Web Mining into three areas of interest based on which part of the Web is used for mining: Web content mining, Web structure Mining and Web Usage Mining [10]. Figure 1 shows the taxonomy of Web mining. Web content mining and Web structure mining utilize the real or primary data on the Web, while Web usage mining mines the secondary data derived from the interactions of the users when they interact with the Web. As a pre processing for Web mining tasks, Web page cleaning mines the inner content of Web pages to discover rules for noise cleaning. Thus, Web page cleaning is a task of Web content mining [11].

2.1. Web Content Mining

Web content mining is the major research area of Web mining. Unlike search engines that simply extract keywords to index Web pages and locate related Web documents for given (keywords based) Web queries, Web content mining is an automatic process that goes beyond keyword extraction [12]. Web content mining directly looks into the inner contents of Web pages to discover interesting information and knowledge. Basically, Web content data consists of texts, images, audios, videos, metadata as well as hyperlinks. However, much of the Web content data is unstructured text data [7] [9]. The research on applying data mining techniques to unstructured text is termed Knowledge Discovery in Texts (KDT) [13], or text data mining, or text mining. According to the data sources used for mining, we can divide Web content mining into two categories: Web page content mining and Web search result mining. Web page content mining directly mines the content of Web pages. Web search result mining aims at improving the search result of some search tools like search engines [14].

2.2. Web Structure Mining

Web structure mining studies the topology of hyperlinks with or without the description of links to discover the model or knowledge underlying the Web [15]. The discovered model can be used to categorize the similarity and relationship between different Web sites. Web structure mining could be used to discover authority Web pages for the subjects (authorities) and overview pages for the subjects that point to many authorities (hubs). Some Web structure mining tasks try to infer Web communities according to the Web topology [16].

Web page cleaning is a crucial pre processing of Web pages for most Web structure mining tasks since the linkages in noisy parts of the Web pages are usually harmful to Web connectivity analysis.

2.3. Web Usage Mining

Web usage mining is the third category in web mining. This type of web mining allows for the collection of Web access information for Web pages. Usage mining also allows companies to produce productive information pertaining to the future of their business function ability. Some of this information can be derived from the collective information of lifetime user value, product cross marketing strategies and promotional campaign effectiveness [17]. Web usage mining is the process of extracting useful information from server logs i.e. users history. Web usage mining is the process

of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data [18].

III. WEB PREFETCHING

Web pre-fetching is another very effective technique, which is utilized to complement the Web caching mechanism. The web pre-fetching predicts the web object expected to be requested in the near future, but these objects are not yet requested by users. Then, the predicted objects are fetched from the origin server and stored in a cache. Thus, the web pre-fetching helps in increasing the cache hits and reducing the user-perceived latency [7].

Web pre-fetching is a technique that made efforts to solve the problem of these access latencies. Specially, global caching methods that straddle across users work quite well. However, the increasing trend of generating dynamic pages in response to HTTP requests from users has rendered them quite ineffective. Pre-fetching is used as an attempt to place data close to the processor before it is required, eliminating as many cache misses as possible. Caching offers the following benefits: Latency reduction, Less Bandwidth consumption, Lessens Web Server load. Pre-fetching is the means to anticipate probable future requests and to fetch the most probable documents, before they are actually requested. It is the speculative 40 retrieval of a resource into a cache in the anticipation that it can be served from the cache in the near future, thereby decreases the load time of the object [8, 19].

IV. TYPES OF WEB PRE-FETCHING

Web Pre-fetching techniques can be divided into two type location based web pre-fetching and link based pre fetching.

- **Location based pre fetching:-** Location based pre fetching can be implemented on server, proxy or client side. The client-based pre-fetching concentrates on the navigation patterns of a single user across many Web servers. On another hand, the sever-based pre-fetching concentrates on the navigation patterns of all users accessing a single website [7]. The proxy-based pre-fetching concentrates on the navigation patterns of a group of users across many Web servers. Thus, this approach can reflect a common interest for user's community. In other words, the pre-fetching contents can be shared by many users [7,8].
- **Link based pre fetching:-**Link pre-fetching is where a web page tells your browser that you are likely to visit a certain page next so your browser should immediately request that next page even though you haven't actually gone there yet.[19] DNS pre-fetching is where your browser tries to speed up future requests by resolving the IP address of every link on web pages you visit (just in case you decide to click on them).

V. RELATED WORK

Some author gives an idea of web pre-fetching. It also discuss about the web object pre-fetching. Some authors describe web mining techniques. They explain how the web mining works. What is the taxonomy of web mining? As the web mining uses in order to pre-fetch the web page [1]. Web mining is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web. Web mining allows you to look for patterns in data through content mining, structure mining, and usage mining. There are other papers presents the idea to use the web log. How the web log can use with other method in different purposes. Here we see that web log is a very important file to investigate the crime. Similarly it can useful to prefetch the web page in efficient manner. A server log is a log file (or several files) automatically created and maintained by a server of activity performed by it [5].

Here the studied papers also throw some light on Markov model. This is a mathematical model based on probability theory. The simplest Markov model is the Markov chain. It models the state of a system with a random variable that changes through time. In this context, the Markov property suggests that the distribution for this variable depends only on the distribution of the previous state. Here we found that it can be apply for the web pre-fetching [20,21].

Various approaches have been developed in improving the efficiency of Web servers, including improved hardware (speed, bandwidth) and software solutions (more suitable models and protocols, better algorithms) [9,10]. A commonly used and effective technique is pre-fetching that preloads some data to the local cache before it is actually requested anticipating that these data are to be requested by the user in the near future so that they will be readily available locally rather than retrieved from remote sites. Of course, the preloading process is to retrieve from remote sources, but it can be done without perceived delay from the user's point of view, simply because there is always a time gap between consecutive requests from the same user in the Web environment and the Web server can use this time gap to pre-fetch the predicted pages [1, 2, 4]. Successful pre-fetching will not only reduce the delays for users' requests for Web objects, but also result in less overall network traffic and lighter loads on the Web servers.

VI. CONCLUSIONS

There are large number of web application has been used for the various purpose. These applications should be good at its response time. Therefore the web caching and web prefetching are the approaches which can be applicable for the enhancing the response time of a web application. This paper is a assessment on these approaches. These concepts come under the web mining so here we also covered the web mining. This paper also throws some light on the previous work as in the related work. In related work we found that the web prefetching can apply on any web application. This approach can apply with various strategies.

In future we would like to apply the web mining as the concept of web prefetching in order to enhance the quality services of a web application with hybrid approach.

ACKNOWLEDGMENT

I would like to say thanks to my guide Dr. Bhupesh Gour and Dr. Asif Ullah Khan who gives their knowledge and time in order to complete this paper. This paper will never complete without the support of faculty members CSE department of TIT College, Bhopal.

REFERENCES

- [1] K.Ramu, Dr.R.Sugumar and B.Shanmugasundaram "A Study on Web Prefetching Techniques" Journal of Advances in *Computational Research: An International Journal* Vol. 1 No. 1-2 (January-December, 2012)
- [2] Waleed Ali, Siti Mariyam Shamsuddin, and Abdul Samad Ismail "A Survey of Web Caching and Prefetching", *Int. J. Advance. Soft Comput. Appl.*, Vol. 3, No. 1, March 2011
- [3] Daesung Lee and Kuinam J. Kim, "A Study on Improving Web Cache Server Performance Using Delayed Caching", *IEEE 2010*, pp 1-5.
- [4] Greg Barish and Katia Obraczka, "World Wide Web Caching: Trends and Techniques", *IEEE Communications Magazine* May 2000
- [5] Pablo Rodriguez, Christian Spanner, and Ernst W. Biersack, "Analysis of Web Caching Architectures: Hierarchical and Distributed Caching", *IEEE/ACM Transactions On Networking*, Vol. 9, No. 4, August 2001
- [6] L. Ramaswamy, A. Iyengar, L. Liu, F. Douglass, "Automatic Fragment Detection in Dynamic Web Pages and Its Impact on Caching", *IEEE Transactions On Knowledge And Data Engineering*, Vol. 17, No. 6, June 2005.
- [7] M. Junchang, G. Zhimin, "Finding Shared Fragments in Large Collection of Web Pages for Fragment-based Web Caching", *Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)* 2006.
- [8] Bhawna Nigamand Dr. Suresh Jain, "Analysis Of Markov Model On Different Web Prefetching And Caching Schemes", *IEEE 2010*, 978-1-4244-5967-4/10/
- [9] P. Kolari and A. Joshi, "Web mining: Research and practice", *Computer Science Engineering* July/August (2004) 42-53
- [10] B. Liu and K. Chang, "Editorial: Special issue on web content mining", *SIGKDD Explorations* 6(2) 2004, pp 1-4.

- [11] Nacim Fateh Chikhi, Bernard Rothenburger, Nathalie Aussenac-Gilles "A Comparison of Dimensionality Reduction Techniques for Web Structure Mining", *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence 2007*, pp 116-119.
- [12] Lefteris Moussiades, Athena Vakali, "Mining the Community Structure of a Web Site," *bci Fourth Balkan Conference in Informatics 2009*, pp.239-244.
- [13] Jaideep Srivastava y, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", *ACM 2000*, PP 12-23.
- [14] Feng Zhang Hhui-You Chang "Research and development in web usage mining system--key issues and proposed solutions: a survey", *Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 4-5 November 2002*
- [15] M. Craven, D. Dipasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery, "Learning to extract symbolic knowledge from the World Wide Web", *In proceeding of the Fifteenth National Conference on Artificial Intelligence (AAAI98)*, pages 509-516, 1998.
- [16] S. Chakrabarti, "Data mining for hypertext: A tutorial survey". *ACM SIGKDD Explorations*, 1(2):1-11, 2000.
- [17] R. Feldman, M. Fresko, Y. Kinar, Y. Lindell, O. Liphstat, M. Rajman, Y. Schler, and O. Zamir, "Text mining at the term level. In Principles of Data Mining and Knowledge Discovery", *Second European Symposium, PKDD'98, volume 1510 of Lecture Notes in Computer Science*, pages 56-64. Springer, 1998.
- [18] S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, S. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Mining the link structure of the World Wide Web", *IEEE Computer*, 32(8):60-67, 1999.
- [19] Brian D. Davison, "A Web Caching Primer" *IEEE Internet Computing 2001*
- [20] Borges and M. Levene, "A dynamic clustering-based markov model for web usage Mining", *cs.IR/0406032*, 2004.
- [21] Zhu, J., Hong, J. and Hughes, J. G. (2002a) Using Markov Chains for Link Prediction in Adaptive Web Sites. *In Proc. of Soft-Ware 2002: the First International Conference on Computing in an Imperfect World*, pp. 60-73, *Lecture Notes in Computer Science*, Springer, Belfast, April.

BIOGRAPHY

Veena Singh Bhadauriya received B.Tech. From UPTU Lucknow in 2006. Pursuing M-Tech from RGPV University, Bhopal.



Bhupesh Gour received the B.E. from GEC, Jabalpur in 2000, M.Tech from RGPV in 2005. He received the Dr. Eng. degree from RGPV University in 2011. Currently he is working as professor CSE at Technocrats Institute of Technology, Bhopal. His research interest includes image processing, Neural Networks and Fuzzy Logic. He is a member of IACSIT, Senior Member IEEE, CSI and LM-ISTE.



Asif Ullah Khan received the B.E. from SATI, Vidisha in 1990 with distinction, Gold Medalist at M.Tech from RGPV in 2005. He received the Doctorate in CSE from RGPV in 2009. Currently he is working as professor CSE at Technocrats Institute of Technology, Bhopal. His research interest includes Image processing, Soft Computing and Data Mining. He is a Senior member of IEEE, Senior Member IACSIT, NM-CSI and LM-ISTE.

