

Improved Prefetching Methodology to Enhance Web Personalization

S.R.Balasundaram¹ S.Akhilan²

¹ Assistant Professor, Department of Computer Applications
blsundar@nitt.edu

² Research Scholar, Department of Computer Science and Engineering
National Institute of Technology, Tiruchirappalli – 620 015
akhilan @ nitt.edu

Abstract – Personalization provides the web pages as per the needs of the web users. Prefetching refers to fetching information from web servers even before they are requested. The prefetching process will be highly essential for the personalization of web details. LRU algorithm is the most commonly available algorithm in the prefetching process. In this paper, the significance of prefetching and caching are discussed and also the importance of LRU-K algorithm is highlighted.

Key words - Personalization, Page rank, Web caching, Prefetching, LRU – K

I. INTRODUCTION

On a Web site, personalization is the process of tailoring pages to individual users' characteristics, preferences and needs. Personalization is used to provide commonly used customer services or e-commerce transactions. Since, the web page is tailored explicitly as per individual requirements; it is also referred as one-to-one marketing. It makes interactions faster and easier. Broadvision, Response Logic, and Autonomy are the well known personalization software products available. The aims of personalization include serving the customer by anticipating needs, making the interaction efficient and satisfying for user and customer. It enables to build a relationship that encourages the customer to do subsequent purchases. Thus, user satisfaction is the ultimate aim of personalization. The technology of personalization includes collaborative filtering, user profiling and data analysis tools. Once the user's needs are established, rules and techniques can be adopted to provide personalized services [1, 2]. Amazon is the best example for providing personalized information.

II. PAGE RANK

When a user wants to view a web page in the web, the user puts a query to search engine. In turn the search engine lists web sites with hundreds and

thousands of web pages as a result of the query put by the user. So the ranking of these web sites and its pages are very important. Page rank is a method for computing the rank of every web page based on the graph/link of the web [5]. Page rank is proposed to measure the relative importance of web pages in the web.

Unlike normal document collections, the World Wide Web is hypertext and provides considerable auxiliary information on top of the text of the web pages, such as link structure and link text. With the help of this link structure in the web, page rank of every web page can be identified. This will help web users and search engines to make sense of the immense or vast heterogeneity of the World Wide Web. The link structure of the Web is much more dynamic than the contents on the Web.

For example the Page rank of page A which has the pages L_1 to L_n is given as follows.

$$PR(A) = (1-d) + d (PR(L_1) / C(L_1) + \dots + PR(L_n) / C(L_n))$$

where

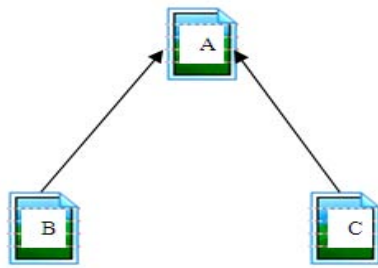
PR(A) - Page Rank of a page A

PR(L₁) - Page Rank of a page L₁

C(L_n) - Number of outgoing links from the page L₁

d - Damping factor ($0 < d < 1$).

Page rank is a global ranking based on the web's graph structure. Also, page rank uses (Fig.1) back links information to bring order to the web.



B and C are back links of A

Fig. 1 Page rank diagram

III. WEB CACHING AND PREFETCHING

A web cache sits between one or more web servers and a client or many clients, and processes the requests and saves the copies of the responses made to the requests. Then, if there is another request for the same URL, it can use the response that it has, instead of asking the web server for it again. Web caches are used to reduce latency and to reduce network traffic. Caching can be done at various levels in a network. It lies on the side of a web server, a caching proxy, or a client. Caching proxies are a particular kind of servers that are responsible for identifying a cached copy of a web page. Caching plays a major role in designing system to get better and fast performance in World Wide Web [3].

Web caching is different from traditional caching in different ways. An important difference is that the size of web objects is not uniform, and the transfer time costs are not identical. Due to the increased web usage and its development a huge amount of network traffic prevails in present web services. It is very much essential to reduce this traffic for utilizing the internet facilities in an efficient manner. Web caching improves the overall performance of web and its related services. At the same time the benefits of caches are minimal. One way to increase the cache benefits is to anticipate future requests and prefetch these web pages into a local cache. This minimizes the user access latency. Web prefetching provides better solution to user latency and bandwidth usage. Prefetching reduces the network traffic that prevails in the modern web services. Prefetching minimizes the user access time but at the same time performance measurement of prefetching techniques are primarily in terms of bandwidth usage [4]. Various prefetching and caching algorithms are in use widely in computing environment. In this paper, a discussion of LRU-K

algorithm and its suitability for caching are discussed.

IV. RELATED WORKS

Victor Safronov and Manish Parashar et.al. (2006) have proposed Page Rank-based prefetching mechanism for web page accesses [9]. They rank the pages linked to a requested page and use the rank to determine the pages to be prefetched. P. Jelenkovi et al. (2008) have proposed a class of randomized LRU caching algorithms which are used to mitigate the effect of variable document sizes on Web caching [6]. Hua Chen et al. (2006) have proposed a new prediction algorithm for deciding upon which page should be pre-fetched [7]. Piyush Kumar et al. (2004) have proposed a model that assumes a cache line chosen for replacement is the one that is accessed furthest in the future [4]. LRU algorithm is the most commonly preferred prefetching techniques available.

V. LRU ALGORITHM

LRU is designed according to the temporal locality of an ordered list of web pages. It is slow on memory reference but fast on page replacement [8]. In a web, a page that is recently accessed can be accessed again in the future. In LRU algorithm, the page is replaced which is not available for the longest time in the cache. At the same time, it is necessary to keep track of the recently viewed web page on every access in LRU algorithm.

VI. LRU – K ALGORITHM

In LRU-K page replacement algorithm, the key idea is to keep on tracking the last K references of the web pages viewed by the user. When the situation arises that the cache is full, LRU-K replaces the page with the leading K length. But, there can be chance of several pages to have the maximum length of the leading web page. In this context, in order to decide which page is to be selected, several links can be considered. LRU- K captures only recently used web pages. Better decisions can be made with the help of recently used pages. By this way, LRU – K can improve the personalization process.

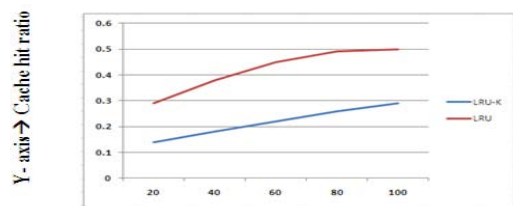
Comparison of LRU and LRU - K algorithms is given in Table 1.

Table 1. Comparison of LRU and LRU – K Algorithms

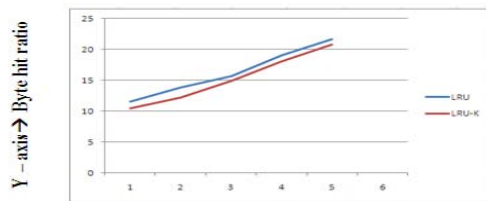
LRU	LRU – K
Tracks the least recently used page	Tracks the last K references of the web page
Designed according to the temporal locality	Designed based on page length
Not more effective in its performance	More effective in its performance
Captures only recently and frequently used web pages.	Captures only recently used web pages.

REFERENCES

- [1] Magdalini Eirinaki and Michalis Vazirgiannis, "Usage-based Page Rank for Web Personalization," Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05), April 2005.
- [2] Bin Hua, Kok Wai Wong and Chun Che Fung, "Fuzzy Logic Based Product Filtering For Web Personalization in E-Commerce," "Proceedings of the IEEE International Region 10 Conference, Melbourne, Australia, 5pp. Nov 2005.
- [3] Evangelos P. Markatos, "On Caching Search Engine Query Results," In Proceedings of the 5th International Web Caching and Content Delivery Workshop, May 2000.
- [4] Piyush Kumar, "Cache Oblivious algorithms," Web Technology and Systems, 28(7-11):893–906, March 2004.
- [5] Victor Safronov and Manish Parashar, "Optimizing Web Servers Using Page Rank Prefetching for Clustered Accesses," In Proceedings of the Conference on Internet Technologies and Systems, pages 193–206, April 2006.
- [6] P. Jelenković and A. Radovanović, "Optimizing the LRU Algorithm for Web Caching," World Wide Web: Internet and Web Information Systems, PP. 143–179, September 2008.
- [7] Hua Chen, "Page replacement algorithm using decision procedure," Proceedings of Sixth International Conference on Web Intelligence and Systems, November 2006.
- [8] Allan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal and Panayiotis Tsaparas, "Link Analysis Ranking: Algorithms, Theory and Experiments," ACM Transactions on Internet Technology, Vol. 5, No. 1, February 2005.
- [9] Victor Safronov and Manish Parashar, "Optimizing Web Servers Using Page Rank Prefetching for Clustered Accesses," The Internet Protocol Journal, vol 7, num 8. September 2006.



(a)



(b)

Fig. 2 (a) Comparative analysis of LRU and LRU-K (No. of pages vs. cache hit ratio)

(b) Comparative analysis of LRU and LRU-K (No. of pages vs. Byte hit ratio)

Comparison between LRU and LRU-K based on No. of pages vs. {Cache hit ratio, Byte hit ratio} is shown in Fig. 2.

VII. CONCLUSION

Personalization is used to customize the content of the web pages automatically for each user. Web caching plays a major role in the personalization process. The performance of caching is improved by prefetching techniques. LRU is the most significant prefetching algorithm. In this paper the significance of LRU – K an extension of LRU is discussed.