Novel Approaches for Integrating MART1 Clustering Based Pre-Fetching Technique with Web Caching

Sathiyamoorthi, CSE, Sona College of Technology, Salem, Tamil Nadu, India
Murali Bhaskaran, Paavai College of Engineering, Nammakal, Tamil Nadu, India

ABSTRACT

Web caching and Web pre-fetching are two important techniques for improving the performance of Web based information retrieval system. These two techniques would complement each other, since Web caching provides temporal locality whereas Web pre-fetching provides spatial locality of Web objects. However, if the web caching and pre-fetching are integrated inefficiently, this might cause increasing the network traffic as well as the Web server load. Conventional policies are most suitable only for memory caching since it involves fixed page size. But when one deals with web caching which involves pages of different size. Hence one need an efficient algorithm that works better in web cache environment. Moreover conventional replacement policies are not suitable in clustering based pre-fetching environment since multiple objects were pre-fetched. Hence, it cannot be handled by conventional algorithms. Therefore, care must be taken while integrating web caching with web pre-fetching technique in order to overcome these limitations. In this paper, novel algorithms have been proposed for integrating web caching with clustering based pre-fetching technique. Here Modified ART1 has been used for clustering based pre-fetching technique. The proposed algorithm outperforms the traditional algorithms in terms of hit rate and number of objects to be pre-fetched. Hence saves bandwidth.

Keywords: Latency, Prefetching, Spatial Locality, Temporal Locality, Web Caching

INTRODUCTION

Web pre-fetching used to improve the performance of web based information retrieval system by pre-fetching web pages into the cache before actual request arrives. Therefore, Web pre-fetching involves following steps

- Prediction system for anticipating user future request based on their previous access pattern.
- Pre-fetching and loading them into the cache.

Thus, web pre-fetching also involves web caching. However, the web caching and web pre-fetching were addressed separately by many
researchers in the past. Only few research works were carried out in integration of web caching and web pre-fetching. The following section gives an overview of all such techniques.

Most of the existing pre-fetching techniques employ single object pre-fetching technique, which is handled by traditional cache replacement policies. However, in the clustering based pre-fetching technique, multiple objects are pre-fetched. Hence, existing replacement algorithms are not suitable for multiple object pre-fetching technique. Hence this chapter provides different heuristics techniques in cache replacement policy which will ultimately reduces the number of objects pre-fetched, increases prediction accuracy and hence saves the bandwidth.

RELATED WORK

More and more researcher has focused on web caching and pre-fetching techniques in recent years (Acharjee, 2006; Huang & Hsu, 2008; Pallis, 2008; and Feng, 2009). It is important to consider the impact of these two techniques when combined together. The authors in Kroeger (1997) have discussed and combined web caching and web pre-fetching to improve delay in accessing web pages. The results shows that combination of web caching and web pre-fetching improved latency up to 60%, whereas web caching alone improved the latency up to 26%. Similarly, the authors in Yang (2001) have proposed pre-fetching technique based on web usage mining. They have used web mining technique called web usage mining to obtain web access patterns and used these patterns to enhance GDSF cache replacement policies with web pre-fetching policy. The authors in Teng (2005) have proposed a cache replacement algorithm called IWCP for integrating Web caching and Web pre-fetching.

Similar to Ibrahim and Xu (2004), the authors in Acharjee (2006) have used ANN in both web pre-fetching and Web cache removal policy. In their approach, keywords present in URL anchor text is used to predict the future request. The main drawback of it is that they have not considered factors such as recency and frequency in web cache removal policy. Also, the keywords extracted from web documents were given as inputs to ANN and hence it might cause extra overhead on the server side.

The authors in Jin (2007) have presented an algorithm for integrating web caching and web pre-fetching technique for wireless local area network. Similarly, the authors in Sulaiman (2009) have presented a framework for integrating Web caching and web pre-fetching on mobile environment.

All the previous works discussed above were used to integrate the web pre-fetching with caching. However, these approaches are still not efficient enough (Liu, 2009). Most previous works were single object pre-fetching technique which uses association rules for pre-fetching approach. They are inaccurate and inefficient since these works predict a particular page depending on patterns observed from all users’ references (Khalil, 2009; Xiao, 2001). Moreover, these approaches employ the traditional replacement algorithms that are inefficient in web caching which involve clustering based pre-fetching technique.

WEB CACHING POLICIES

Web caching policy decides which pages to be removed from the cache when there is no enough space for storing new page. The main goals of web caching are

- It has to use the cache space more efficiently,
- It has to optimize one or several web cache performance metrics such as cache hit rate, byte hit rate, response time and so on.
- It has to improve CPU and memory utilization.

If the future requests are known then web cache system can be further optimized. But the problem is that knowing future requests is obviously impossible. Hence, web cache
A Novel Architecture for Deep Web Crawler
www.igi-global.com/article/novel-architecture-deep-web-crawler/52804?camid=4v1a

A Context-Based Approach to Web 2.0 and Language Education
www.igi-global.com/chapter/context-based-approach-web-language/37681?camid=4v1a