Chapter V

Interactive Proxy for URL Correction

Kai-Hsiang Yang
National Taiwan University, R.O.C.

Abstract

This chapter will address the issues of Uniform Resource Locator (URL) correction techniques in proxy servers. The proxy servers are more and more important in the World Wide Web (WWW), and they provide Web page caches for browsing the Web pages quickly, and also reduce unnecessary network traffic. Traditional proxy servers use the URL to identify their cache, and it is a cache-miss when the request URL is non-existent in its caches. However, for general users, there must be some regularity and scope in browsing the Web. It would be very convenient for users when they do not need to enter the whole long URL, or if they still could see the Web content even though they forgot some part of the URL, especially for those personal favorite Web sites. We will introduce one URL correction mechanism into the personal proxy server to achieve this goal.
Introduction

With the rapid expansion of the World Wide Web (WWW), more and more Web-based applications have caused serious performance degradation on the Internet. Caching is the process of storing Web elements (pages, files, images) on proxy servers. The use of proxy server makes serving information on the Internet more efficient in terms of reducing bandwidth costs and server utilization. The Internet Caching Resource Center (www.caching.com) estimates that caching can reduce the need for bandwidth by at least 35%. Consequently, the proxy servers have been widely deployed to reduce the bandwidth for the same Web page requests; they could accelerate the browsing speed by storing current Web pages for future requests. Nowadays, proxy servers are necessary for the WWW community.

In the traditional proxy server, when it receives a request URL, first it has to check the validation of the URL by the Domain Name System (DNS) lookup. Then the request URL is matched with all URLs in its caches. The proxy server will immediately send back the requested page from its cache if found. Otherwise it has to directly connect to the original server to get the requested page, send it back, and store it in its cache database at the same time. The general proxy server mainly depends on the request URL to operate.

However, this is inconvenient and insufficient for users, especially when they type in the incorrect URL. For example, the correct URL of the Starbucks company is http://www.starbucks.com. We would just get some error messages from the proxy server for the wrong URL (www.starbuck.com), unless we correct it by ourselves.

However for general users, there must be some regularity and scope in browsing the Web, for example, the portals or news Web sites that users browse every day, or some specific Web-based systems. It would be very convenient for users when they enter one word, such as “google” or even with some error, “goggle,” and then would eventually be taken to www.google.com. Therefore, it is a very important and useful function to support the URL correction mechanism based on the personal browsing access logs.

On the other hand, online security has drawn increasing attention, especially for home users. There are several personal proxy tools supporting the functions to block cookies, Web bugs, Web referrers, ads and scripts and stop them from collecting online information. Almost all these tools save the personal access logs into text files. However, they do not utilize these logs to correct URLs.
Grounding CSCW in Social Psychology
www.igi-global.com/chapter/grounding-cscw-social-psychology/13131?camid=4v1a