

Title	先読み代理サーバを用いたWWW情報探索支援
Author(s)	新井, 孝之
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1523
Rights	
Description	Supervisor: 白井 清昭, 情報科学研究科, 修士

Supporting WWW Navigation using Prefetching Proxy Server

Takayuki Arai

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2002

Keywords: Web Navigation, Prefetching Proxy Server, Summarization, User Profile.

The World Wide Web is developing very fast. Currently, finding useful information on the Web often need a lot of time. Usually, one web page have many hyperlinks. When users of web browser search web pages, they select one of hyperlinks, jump other pages, and sometimes return to the original page. However, this operation is very troublesome. Therefore, reducing the operation of jumping other web pages is considered to be useful to gain the efficiency of web searching. In this paper, I propose a system that helps users to effectively browse and search the web pages. When a user puts a mouse cursor on the link on a web page, the system displays the summary of the linked web page. A user can judge whether the linked page is useful or not before jumping and seeing it. In recent years, many researchers have been devoted on an automatic summarization of text, because the amount of the available machine readable text is drastically grown in the World Wide Web. This paper aims at using such summarization techniques for supporting WWW navigation. The system was implemented as a prefetching proxy server, which can be used in any kinds of platforms. The system prefetchs linked web pages, summarizes them and and saves the summaries in a cache before users request the browser to show the summaries of linked pages. However, as a web page has much hyperlinks, prefetching all of the linked pages in a web page simultaneously is not desirable because it causes heavy network traffic. For such a reason, this paper proposed the method using anchor texts and user profile to decide the order of linked web pages in order to prefetch and summarize them sequentially. Therefore, a system can display the summaries of the linked pages quickly to users of web browsers.

An overview of the proposed WWW navigation support system is the following:

1. A user browse any web pages as same as the usual web searching. The system analyzes the web page the user currently sees, and adds Java Script so that a summary of a linked page can be displayed. Furthermore, for all the hyperlinks that exist in the page, it prefetchs the linked web page, summarizes them and saves summaries into a cache. If a user requests the browser to show the summary of a linked page, it immediately extracts them from a cache.
2. If a user places a mouse pointer on a favorite anchor, Java Script will display the summary of a linked web page either in a tool tip near the anchor or in a sub window. By showing a summary, a user can judge whether the linked page is useful or not.

The operations, that a user return to the original page because the linked page is not useful, is expected to be reduced.

3. If a user leave a mouse pointer from an anchor, Java Script will erase a tool tip.
4. If a user actually clicks a link and requests to display the web page, the linked page will be extracted from a cache immediately. If the linked page is not saved into a cache, it acquires it from the Internet.

While a user trolls web pages, the above procedures are repeated. In this paper, a tool tip and a sub window are used as the method to show summaries.

When summarizing the web pages, using techniques based of HTML tags or positions of fragments in a web page seems not to be valid. So this paper uses the existing text summarization techniques, i.e. sentences in a web page are scored based on term frequencies (TF), and the most highly-scored sentences are extracted as a summary.

Prefetching the linked web pages may be cause a heavy network traffic. Therefore, prefetching all of the linked pages in a web page simultaneously is not desirable. The number of prefetched web pages simultaneously should be restricted. The system determine the order of linked pages to prefetch and summarize them sequentially. When determining the order of prefetching, user profile, user's interest in other words, and anchor text are used. More words interesting for a user are contained in the anchor text, the more a user wants to see the the page linked by the anchor. The system saves the words in the web pages which a user browsed in a past, and their frequencies as a user profile. For all words in each anchor text, the system calculates the sum of the frequencies in a user profile as a score, then lists the anchors in order of this score. This is the order of linked pages for prefetching and summarization.

I have implemented the prototype system for supporting WWW navigation, and conducted the experiment for evaluation. As a result, the following knowledge about the system were obtained.

- Showing the summary of a linked page is effective for supporting WWW navigation.
- Prefetching the linked pages was somewhat effective to reduce the time to show summaries.
- However, in the events of displaying summaries of the linked pages, the ratio that the summaries were saved in a cache is 27.6%. Therefore, there is much room to improve the prototype system.
- In the experiment, two method of showing summaries, tool tip and sub window, were compared. As many researches have been used tool tip as a method of showing information in web browsing, tool tip is expected to be useful. However, in this experiment, more user preferred the sub window to tool tip.