An Empirical Analysis of Web Page Revisitation

Bruce McKenzie and Andy Cockburn
Department of Computer Science
University of Canterbury
Christchurch, New Zealand
{bruce,andy}@cosc.canterbury.ac.nz

Abstract

There is a surprising lack of empirical research into user interaction with the web. This paper reports the results of an analysis of four months of logged data describing web use. The results update and extend earlier studies carried out in 1994 and 1995. We found that web page revisitation is a much more prevalent activity than previously reported (approximately 80% of pages have been previously visited by the user), that most pages are visited for a surprisingly short period of time, and that users maintain large (and possibly overwhelming) bookmark collections.

1. Introduction

The World Wide Web, and the web-browsers used to access it, are inextricably linked with most people’s computing experience. Given the predominance of the WWW in everyday computing, there is a surprising lack of research into how the web is used.

Two prior studies provide the empirical foundation for our current understanding of user interaction with the web: Catledge & Pitkow [5] and Tauscher & Greenberg [16]. Though excellent studies, there are several reasons for suspecting that their findings may no longer reflect current use of the web:

1. Age of the studies. The studies were carried out in 1994 (Catledge & Pitkow [5]) and 1995 (Tauscher & Greenberg [16]). Given the relative youth of the web at this time, and its continued exponential growth, it seems reasonable to suspect that usage patterns may have evolved and matured.

2. Web-browser studied. Both studies analysed use of NCSA’s XMosiac browser. Netscape Navigator and Microsoft Internet Explorer are now the dominant web-browsers. Netscape had an estimated 45% share of web-browser use in 1998, and it has been projected that Microsoft Internet Explorer will have a 65% share by 2001 [14]. The user interfaces to the current generation of web browsers have gone through several iterative refinements, and have been the topic of research-level scrutiny: for example, see [2]. It is reasonable to suspect that the improved interfaces may have changed browser usage.

3. Browsers of preference. In [16], Tauscher & Greenberg state that none of the subjects in their study used XMosaic as their normal web-browser. Similarly in [5], Catledge & Pitkow indicate that subjects may have chosen to use a browser other than their specially equipped version of XMosaic. Clearly, the subjects’ behaviour could have been influenced by the use of a non-favoured browser.

4. Duration of the evaluation. Catledge & Pitkow analysed three weeks of user interaction logs with XMosaic, and Tauscher & Greenberg analysed between five and six weeks. It is possible that long term web-page revisitation patterns will be missed in even these fairly long term analyses.

This paper aims to update and overcome some of the limitations of prior empirical investigations into how the web is used. The study presents the results of an analysis of four months of daily client-side log files. The files showed the pages each user visited, the number of times they visited