# WebinSitu: A Comparative Analysis of Blind and Sighted Browsing Behavior

Jeffrey P. Bigham, Anna C. Cavender, Jeremy T. Brudvik, Jacob O. Wobbrock† and Richard E. Ladner

Computer Science and Engineering
Box 352350
University of Washington
Seattle, WA 98195 USA
{jbigham,cavender,jbrudvik,ladner}@cs.washington.edu

†The Information School
Box 352840
University of Washington
Seattle, WA 98195 USA
wobbrock@u.washington.edu

## ABSTRACT

Web browsing is inefficient for blind web users because of persistent accessibility problems, but the extent of these problems and their practical effects from the perspective of the user has not been sufficiently examined. We conducted a study *in situ* to investigate the accessibility of the web as experienced by web users. This remote study used an advanced web proxy that leverages AJAX technology to record both the pages viewed and the actions taken by users on the web pages that they visited. Our study was conducted remotely over the period of one week, and our participants used the assistive technology and software to which they were already accustomed and had already configured according to preference. These advantages allowed us to aggregate observations of many users and to explore the practical effects on and coping strategies employed by our blind participants. Our study reflects web accessibility from the perspective of web users and describes quantitative differences in the browsing behavior of blind and sighted web users.

## Categories and Subject Descriptors

K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

## General Terms

Human Factors, Experimentation

## Keywords

web studies, web accessibility, blind users

## 1. INTRODUCTION

Browsing the web is inefficient for blind web users because of persistent accessibility problems. However, the extent of these problems and their practical effects on browsing experience are not yet adequately understood from the perspective of blind web users. For web access guidelines, standards, and future improvements to be truly relevant and useful, more information about real-life web interaction is needed. In this work, we seek to understand the accessibility of the web from the user perspective by measuring the accessibility of the pages that users visit and comparing the behavior observed in blind users to their sighted counterparts.

We used an advanced web proxy to enable our study and quantitatively measured both the presence and observed effectiveness of components thought to impact web accessibility. Most proxy systems can only record HTTP requests and cannot easily discern user actions performed on web pages [8, 20]. We used an enhanced version of UsaProxy [5] to record participants' browsing. UsaProxy can record actions that are impossible to record with a traditional proxy, such key presses, clicks on arbitrary page elements (including within-page anchor links), and the use of the "back" button to return to a page that was previously viewed. Recording user actions has traditionally required study participants to install specialized browser plugins [15, 9], but UsaProxy is able to record most user actions by using Javascript code that is injected into pages that are viewed. Because it uses Javascript to parse the viewed web pages, it can also record dynamic page changes, interaction with dynamic content and AJAX requests, which are an increasingly important accessibility concern. A proxy approach enables transparent setup by participants and allows them to use their own equipment with its existing configuration.

Prior work has sought a better understanding of the web user experience [20, 22]. The importance of measuring accessibility *in situ* from the user perspective is illustrated by the relative popularity of web sites visited by web users in our study, as shown in Figure 1. The distribution is Zipf-like [7], which results in three sites (google.com, myspace.com and msn.com) accounting for approximately 20% of the pages viewed by the participants in our study. The google.com domain alone accounted for almost twice as many page views as the 630 domains that were viewed five or less times during our study. The accessibility of popular sites more strongly affects users than do sites on the long tail of popularity. While our study is not a replacement for laboratory studies that use common tasks, it offers an important view of accessibility that better matches the experiences of real users.

Blind web users have proven adept at overcoming accessibility problems, and one of the goals of this study was to


*ASSETS'07,* October 15–17, 2007, Tempe, Arizona, USA..

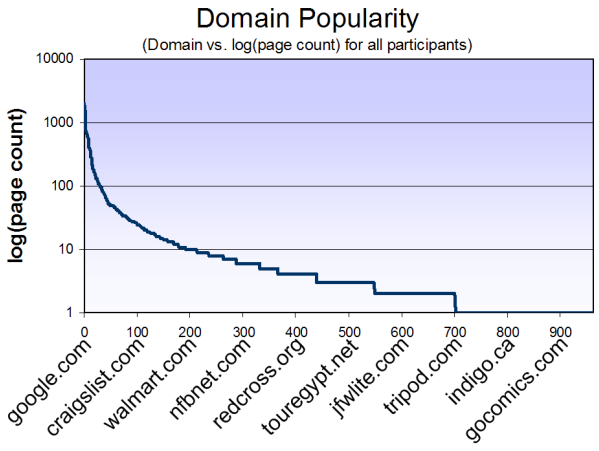


**Figure 1: Log frequency of visits per domain name recorded for all participants ordered by popularity.**

better understand the coping strategies employed by blind users as they browse the web. For instance, the lack of alternative text is an often-cited accessibility concern, but blind users can often obtain the same information contained within an image from surrounding context. Within-page anchors called "skip links" are designed to help blind users effectively navigate complex web pages by enabling them to jump to relevant content, but these links may be used infrequently because other screen reader functionality also enables users to move non-linearly through a page. If the context surrounding links on a page isn't clearly expressed to blind users, they may explore the page by clicking on links simply to see where they point and then return. WebinSitu explores whether blind web users avoid inaccessible content and also if they make use of coping strategies.

The direct effects of technology and developer practices for improving accessibility are difficult to measure in practice because users employ many different browsing and coping strategies that may vary based on the user's familiarity with the page be accessed. Related work has looked at task-based analysis of accessibility [24, 27, 10, 25], with a major focus on supporting effective accessibility evaluation (see Ivory for a survey of this work [21]). Realistic studies with blind web users are difficult to conduct in the lab due to difficulties in replicating the diversity of assistive technology and configurations normally used by participants. Previous work has advocated remote studies because they allow participants to use their existing assistive technology and software [24, 25, 16]. These studies noted that blind participants can be ineffective at providing feedback when a page is considerably inaccessible, indicating that simply asking blind users to list the problems they face may not be sufficient.

Overall, we found that blind web users browse the web quite similarly to sighted users and that most pages visited during our study were inaccessible to some degree. In our study these problems are placed in the context of their predicted effects because we implicitly weighted pages relative to their popularity. Perhaps most surprising, blind participants generally did not shy away from pages exhibiting accessibility problems anymore than did sighted users. Blind participants were, however, much less likely to visit pages containing content not well addressed by assistive technology. Blind users tended not to visit sites heavily dependent on Asychronous Javascript and XML (AJAX), but visited many pages that included Flash content. Blind users also interacted less with both dynamic content and inaccessible web images. Skip links, added to web pages to assist screen reader users, were only used occasionally by our participants. Our analysis highlighted several areas that may suggest the coping strategies used by blind web users when faced with inaccessible content.

The contributions of this paper are as follows: 1) We report on web accessibility as experienced by real web users. 2) We compare the browsing experience of sighted and blind web users on several quantitative dimensions. 3) We demonstrate the effectiveness of proxy-based recording for exploring the interaction of blind web users. 4) We formulate practical user observations that can influence the direction of future web accessibility research.

## 2. RECORDING DATA

We used a tracking proxy to record statistics about the web experience of our participants (see the diagram in Figure 2). The proxy is an extended implementation of UsaProxy, which allows both HTTP request data and user-level events to be recorded [5]. This method of data collection allows participants to be located remotely and use their own equipment. This is important for our study because of the diversity of assistive technology and configurations employed by blind users. Our proxy-based approach requires minimal configuration by the user and does not require the installation of new software. Connecting to the system involved configuring their browsers to communicate with the tracking proxy and entering their login and password. Names and passwords were not connected with individuals, but a record was kept indicating whether the participant primarily uses a screen reader or a visual browser to browse the web.

A browsing session begins with the participant initiating an HTTP request, which is first sent to the proxy and then passed directly to the web server. The web server sends a response back to the proxy, which logs statistics about the response header and web page contents. The proxy also injects JavaScript into HTML responses to record user-generated events and sends this modified response back to the user. After the response is received by the user and is loaded in their browser, the Javascript inserted into the page can record events such as key presses, mouse events, and focus events and sends data about each event, including the Document Object Model (DOM) elements associated with each event, back to the proxy for logging. For example, if a user clicks on a linked image, the click event and its associated image (dimension, source, alternative text, etc.), the link address and position in the DOM are sent to the proxy and recorded. The proxy also records whether content with which participants interact is dynamic (i.e. created after the page was loaded via Javascript) and whether the pages viewed issue AJAX requests.

All of the data pertaining to a participant's browsing experience is stored on a remote database. At any time during the study, participants may examine their generated web traces, comment on the web pages viewed, enter general comments about their browsing experience or delete portions of their recorded browsing history (See Figure 6). Our participants deleted only three browsing history entries.