WebNet 96

Proceedings of *WebNet 96*—

World Conference of the Web Society

San Francisco, California, USA; October 15-19, 1996

AACE Association for the Advancement of Computing in Education

Published by

Association for the Advancement of Computing in Education (AACE)
P.O. Box 2966
Charlottesville, VA 22902 USA

Printed in the USA

FULL PAPERS

Tracking WWW Users: Experience from the Design of HyperVisVR

Chris Brown
Communications Research Group
University of Nottingham, University Park
Nottingham NG7 2RD ENGLAND
ccb@cs.nott.ac.uk,

Steve Benford
Communications Research Group
University of Nottingham, University Park
Nottingham NG7 2RD ENGLAND
sdb@cs.nott.ac.uk

Abstract: We are generally concerned with the development of a hypertext database visualization framework, HyperVisVR, which supports 3D collaborative visualization, dynamic responsiveness to database updates and representation of the underlying database users in the visualization. In developing a plug-in database access module for the World Wide Web (WWW), we have needed to solve several problems related to the WWW's loosely-coupled stateless architecture. One such problem is that of tracking WWW users as they move between pages and servers. This paper discusses current approaches to tracking WWW users, proposes new ones and explores issues of privacy and mutuality with respect to the monitoring of hypertext access.

Introduction

As the WWW increases in size and complexity, there becomes a need for advanced tools to manage information overload. Image maps, navigation bars, server push/pull, Java, ActiveX and Shockwave are among the layout techniques that are being used by content providers to generate the site designs which predominate the WWW today. Although these complex web site layouts may improve navigability within the site itself, when we consider the WWW as a global distributed hypertext database, they can cause confusion as browsing users are met with a bewildering array of non-standard navigational tools and controls. Despite this recent proliferation of these non-standard interfaces, hypertext navigability is not a new problem. In 1990, Jacob Nielsen addressed the issue of users becoming lost in hyperspace and proposed the use of overview diagrams (maps) and fish-eye techniques [Nielsen 1990]. These approaches have been adopted by several researchers trying to improve the navigability of the WWW, particularly Dmel with his WebMap tool [Domel 1994], and Mujherta with the Navigation View Builder [Mujherta 1994]. The motivation for this paper is a project entitled HyperVisVR, which is a framework for visualization of large hypermedia databases within a 3D virtual world. The driving goals of this framework are:

- Extensible through plug-in modules to support new databases and visualization styles.
- Fully dynamic visualization which is responsive to viewers movement through virtual world.
- Database access is represented in the visualization: users moving through the database have `embodiments'.
- Modifications to the database are reflected in the visualization.
- HyperVisVR applications support peer-peer communications to represent other visualization viewers in the visualization itself, and to share cached information about the database objects.

As an underlying database for HyperVisVR, the WWW has the benefit of being ubiquitously available throughout the world. However, we believe it poses some unique problems due to its loosely-coupled, stateless, architecture. This makes it a particularly interesting first target for our visualization framework. To be more specific, the problems with the WWW as a database for HyperVisVR are:

- It is difficult to extract rich meta information about objects on the WWW.

- Objects and links are stored together, making it difficult to analyse the structure of the WWW without performing an exhaustive search through the objects themselves.
- As a connectionless, stateless protocol, HTTP makes it very difficult to track users and their actions on the database on a global scale.

However, we don't believe that these problems are limited to HyperVisVR and many researchers are trying to find ways around the limitations on HTTP as a hypertext database protocol, and HTML as a hypertext object description language. Finding backwardly-compatible solutions to these problems would be of benefit to the research community as a whole. In this paper, we turn our attention to the specific problem of tracking users of the WWW as they browse and search the global WWW infrastructure. Potential applications of our ideas include:

- WWW providers mightoptimise their servers byanalysing the paths of users through their servers.
- WWW providers might provide their clients with more detailed usage logs.
- WWW applications might provide a mutual awareness of users leading to opportunities for encounters.
- Tools to help users understand the relation between their current location and its surrounding context.
- Search tools might use the activity of users around a set of WWW resources to indicate interest in them.

User Tracking

We now focus on the problem of tracking users as they pass through the WWW. Users can be tracked at three different places. A server can track users moving through it, a browser can be modified to provide tracking information whenever a database access is made and proxy servers can track the users who use the proxy to access other areas of the WWW. The first two sub-sections below assume that all connections are direct between the user agent and the origin server. Proxies further complicate user tracking techniques, and are considered in section 2.3

Tracking at the Server

The key reason that user tracking at the server side is so difficult is that HTTP is stateless. Each request for an object from an HTTP server is completely separate, and cannot easily be associated with previous requests from the same user. The following sections discuss methods for user tracking that are available with current and proposed versions of HTTP, extensions to browsers that enable user tracking, and suggestions for improvements to HTTP which would make user tracking cleaner and more reliable.

In HTTP/1.0

There are several request headers specified in HTTP/1.0 [Berners-Lee et al. 1996] which can be used by the server to link up requests into a click-trail for a particular user in a given session. These are: From and Referer[sic]. In addition, the IP address of the machine generating the request can be used as a basis for tracking. The From request header is an optional field, specified to contain the Internet e-mail address of the user. However, only several less popular browsers actually make use of this field, as it is regarded by most to be a breach of the privacy of the user to send his personal information in each request. From fields are most widely used in WWW robots to identify the administrator in case there is a problem with the robot. The Referer request header can also help track click paths. This field can be used by the browser to specify the address of the resource from which the current request address was obtained. This can be interpreted by the server to link a previous request to the current one.

As browsers follow outside links from the server to other servers, their movement can be tracked through the use of cgi-redirect scripts. All outside links on a server are modified so that they link to a local cgi-redirect script, and pass the location of the outside link to the script as a parameter. The cgi-redirect script can log the movement, and then issue a redirect to automatically route the browser to the outside resource.

Some sites use intelligent algorithms which analyse log files and link up requests into click-trails through the server using a combination of IP address, the From and Referer request headers (if present), a maximum time

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.