

Transient User Profiling

Ed H. Chi

Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304 USA
echi@parc.com

Abstract

Our work in the past five years on modeling user actions on the Web has shown that a great deal of information about user actions can be recovered from the informational cues processed by the user during navigation. We call these informational cues by the name of “Information Scent.” We have shown in various papers that Information Scent can be used as a methodology for clustering a group of user profiles [Chi02], simulating a collection of users navigating thru the Web with an information need [Chi03], and providing navigational cues to users with transient information goals [Olston03].

We argue in this position paper for a CHI2004 workshop that more research in user profiling should be done for user goals that are transient in nature.

Keywords

User Profiling, Transient Information Goals, Information Scent, User Modeling

MOTIVATION

1. When do profiles have an added value?
Much research has gone into understanding how to store and understand long-term user preferences on things such as informational topics, television program watching habits (e.g. Tivo), and movies (e.g. movielens.umn.edu). These long-term user profiles have greatly enhanced user interaction with informational sources. For instance, it is of great comfort to an average consumer that Amazon.com remembers her purchase of the DVD release of the first volume of the “Lord of the Rings” trilogy, and suggests subsequent volumes as they are released. As another example of the impact of user profiling systems, a movie recommendation service, movielens.umn.edu, has reported a user who mistakenly believed that the system had thrown away all of his ratings of movies and re-entered all thousand-plus ratings one by one under a new user id. Because they provide such dramatic benefit to the user, long-term user profiles are now a standard tool in eCommerce as well as in other informational environments.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2004

Copyright 2004 ACM 1-58113-453-3/02/0004...\$5.00.

Much more work remains in popularizing long-term user profiles in large information environments. For instance, most current commercially available digital libraries offer little or no long-term analysis of user access patterns or user topic preferences. There are a number of reasons for this:

- First, in order to provide a personalized experience, significant infrastructure must be built and maintained. A large server is often required to store and maintain the large number of user profiles.
- Second, open-source or cheap alternatives to custom-built commercial alternatives do not currently exist.
- Third, there are few experts available to guide through the implementation process.

By nature, long-term user profiling and analysis is a heavy-weight process that requires specific expertise and knowledge to implement correctly.

In addition to these problems, long-term user profiles have also made some well-known blunders. The most often cited example is a male customer buying a copy of a pregnancy book as a gift on Amazon.com only to be offered many more pregnancy books on subsequent visits for months to come. This example stems from the fact that many user interests are ephemeral and transient, and it is extremely difficult for current long-term user profiling methods to distinguish the context in which a user decision is made.

OUR PREVIOUS WORK

Three pieces of our previous work utilize user profiles¹. These works all make use of a concept called Information Scent. We are interested in the frequent case where users have some *information goal*—some specific information they are seeking. The content of Web pages is usually presented to the user by some snippets of text or graphic called *browsing cues*. Foragers use these browsing cues to access the *distal content*: the page at the other end of the link. *Information scent* is the imperfect, subjective perception of the value, cost, or access path of information sources obtained from browsing cues.

- ScentTrails: Integrating Browsing and Searching on the Web [Olston03]. The two predominant paradigms for finding information on the Web are browsing and keyword searching. While they exhibit complementary ad-

¹ These papers can be found at <http://www.geekbiker.com>

vantages, neither paradigm alone is adequate for complex information goals that lend themselves partially to browsing and partially to searching. To integrate browsing and searching smoothly into a single interface, we introduce a novel approach called ScentTrails. Based on the concept of information scent developed in the context of information foraging theory, ScentTrails highlights hyperlinks to indicate paths to search results. In ScentTrails, the user enters key-words representing the portion of his or her information goal that is initially known and amenable to search. ScentTrails then annotates the hyperlinks of Web pages with search cues: indications that a link leads to content that matches the search query. This annotation is done by visually highlighting links to complement the browsing cues already embedded in each page. The degree to which links are highlighted is determined by an Information Scent algorithm. By considering both the search and browsing cues together, the user is able to make informed navigational decisions and efficiently locate content matching complex information goals that lend themselves partially to searching and partially to browsing. This interface enables users to interpolate smoothly between searching and browsing to locate content matching complex information goals effectively. In a preliminary user study, ScentTrails enabled subjects to find information more quickly than by either searching or browsing alone.

- LumberJack: Intelligent Discovery and Analysis of Web User Traffic Composition [Chi02]. Web Usage Mining enables new understanding of user goals on the Web. This understanding has broad applications, and traditional mining techniques such as association rules have been used in business applications. We have developed an automated method to directly infer the major groupings of user traffic on a Web site [Heer01]. We do this by utilizing multiple data features of user sessions in a clustering analysis. We have performed an extensive, systematic evaluation of the proposed approach, and have discovered that certain clustering schemes can achieve categorization accuracies as high as 99% [Heer02b]. We described the further development of this work into a prototype service for clustering user profiles called LumberJack, a push-button analysis system that is both more automated and accurate than past systems.
- The Bloodhound Project: Automating Discovery of Web Usability Issues using the InfoScent™ Simulator. According to usability experts, the top user issue for Web sites is difficult navigation. We have been developing automated usability tools for several years, and here we describe a prototype service called InfoScent™ Bloodhound Simulator, a push-button navigation analysis system, which given a transient user information goal automatically analyzes the information cues on a Web site to produce a usability report. We presented a user study involving 244 subjects over 1385 user sessions that shows how Bloodhound user profiles correlates with real

users surfing for information on four Web sites. The hope is that, by using a simulation of user surfing behavior, we can reduce the need for human labor during usability testing, thus dramatically lowering testing costs, and ultimately improving user experience. The Bloodhound Project is unique in that we apply a concrete HCI theory directly to a real-world problem.

SUGGESTED APPROACH

2. *What should a profile contain, how much intelligence should it possess?*

We believe that more research is needed to understand the relationship between long-term user interests and transient short-term user interests. Specifically, we propose several research directions:

- Transient user interests should be used more often to enhance user interaction. In our research, we asked users to perform tasks on the Web, researching specific products on a site or locating a specific piece of information. Our ScentTrails work [Olston03] shows that by simply asking the user for a few keywords that describe user's current goal, we can enhance her performance by 50% in locating specific information on a Web site. These informational goals are specified in a transient way. One user session do not convey user preferences to the next session. However, our work shows that transient user profiles are extremely effective in helping users find information.
- We propose that there are ways in which transient user interests can be detected, for example, by comparing long-term user profiles with the current information that the user is accessing. A user profile might suggest that when a user searches for "Java", he typically mean the programming language. But for the one instance that he meant the island "Java", it can be detected by the fact that he had just been searching for travel information to the Asia region.
- We propose to enable application designers to dial back and forth between placing greater weight on long-term or transient user profiles. We surmise that, in certain applications, long-term profiles are more effective than transient profiles and vice versa. For instance, in an intranet search application, the word "Java" might always mean the programming language for a software programmer, but always mean the island "Java" for a travel agent. So we surmise that in an intranet application, long-term profiles might be given more weight.

3. *Implementation issues: multiple profiles, ownership, security etc.*

A frustrating experience is that current user profiles do not transfer easily between different computing systems. My user profile at work (such as what I have seen on the Web and the documents that I've worked on) do not transfer easily to the home office computer. With the increasing blurring of the boundary between work and home, knowl-

edge workers are demanding more and more that their working environment is mobile and can be securely transferred between computing systems and services. For instance, what books I have purchased on Amazon.com could inform Google the kind of topics I am interested in.

Many of these profile ownership issues are legal and political issues. Does Amazon.com own my user profile, or do I? Do I control the visibility of my profile or do Amazon?

We believe in a liberal position in that users own the profiles that they create on these services, and services have an obligation to provide ways to access, delete, modify these user profile as well as controlling the visibility and security of the user profiles.

CONCLUSION

4. What is the relationship between profiling and HCI? Examples of profiles and/or (future) scenarios in which profiles play a crucial role across environments are appreciated.

As our own work suggests, user profiles can be used in many different situation in the field of HCI. We have used user profiles to simulate how real users with the information interests in those user profiles might navigate a Web site in the Bloodhound Project. In the ScentTrails project, we showed that this simulation can be used to predict where the users might go, and these predictions can be used to visually augment the interface so that users are more efficient in their navigation. Lastly, in the LumberJack Project, we showed how user profiles can be mined and clustered for understanding of the major interests of groups of users visiting a Web site. This clustering technique can be used to develop better user profiles and used to bootstrap future users that come to the Web site.

From this example, we can see that the concept of Information Scent has been used as a unifying approach to solving many of the user profile issues we are interested in. Information Scent has also been used to construct and build profiles across several applications. We are interested in how understanding the Information Scent approach might help in transferring user profiles in a wide-variety of HCI systems, including mobile information environments, digital libraries, and electronic books. In fact, we have begun to apply Information Scent concepts to help users find information in electronic books in a prototype project. However, we still are seeking answers of how user profiles can be transferred between various applications so that users might have a unifying identity and profile across several information environments. For example, we are interested in how the use of the ScentTrails enhanced browsing system could inform the searches at Google or the reading activity of an electronic book.

User profiling appears to play a significant role in better and more efficient interfaces in the future. The question is not whether it will be used to build future HCI systems, but rather how it will be used.

ACKNOWLEDGMENTS

The works described in this position paper is the result of collaboration with a large number of co-workers at PARC. They are too numerous to list here, but I wish to acknowledge and thank them for their collaboration. This work was supported in part by Office of Naval Research grant No. N00014-96-C-0097 and #MDA904-03-C-0404 from the Advanced Research and Development Activity, Novel Intelligence from Massive Data program to Peter Pirolli and Stuart Card.

REFERENCES

- [1] Ed H. Chi, Adam S. Rosien, Jeffrey Heer. Lumber-Jack: Intelligent Discovery and Analysis of Web User Traffic Composition. In *Proc. ACM-SIGKDD Workshop on Web Mining for Usage Patterns and User Profiles (WebKDD 2002)*. ACM Press, July 2002. Edmonton, Canada.
- [2] Ed H. Chi, Adam Rosien, Gesara Suppattanasiri, Amanda Williams, Christiaan Royer, Celia Chow, Erica Robles, Brinda Dalal, Julie Chen, Steve Cousins. The Bloodhound Project: Automating Discovery of Web Usability Issues using the InfoScent(tm) Simulator. In *Proc. of ACM CHI 2003 Conference on Human Factors in Computing Systems*. ACM Press, April 2003. Fort Lauderdale, FL.
- [3] Heer, J. and Chi, E.H. Separating the Swarm: Categorization Methods for User Access Sessions on the Web. In *Proc. of ACM CHI 2002 Conference on Human Factors in Computing Systems*, pp. 243--250. ACM Press, April 2002. Minneapolis, MN.
- [4] Chris Olston, Ed H. Chi. ScentTrails: Integrating Browsing and Searching on the Web. *ACM Transactions on Computer-Human Interaction*, Vol. 10, Part 3, pp. 177--197. Sept, 2003. ACM Press.

Workshop information:

http://chi2004.org/program/prog_workshops.html#ws4

WS#4. User Profiling

Johan Schuurmans, IBM, The Netherlands
Boris de Ruyter, Philips, The Netherlands
Harry van Vliet, Telematica Instituut, The Netherlands

Some global trends have a large influence on the way we use technology, like:

- increasing connectivity and connected devices
- broadband
- increasing data storage capacity, both local and network
- smart objects (ambient intelligence, intelligent multi-modal user interfaces)
- heterogeneous environments

To transform these trends into benefits for end users, products and services need to make efficient use of user profiles. A possible concern for end users is the large number and wide variety of different profiles to deal with.

In this workshop we will discuss the width and depth of this topic and will try to:

- get more grasp on “the big picture”

- come to conclusions & recommendations in the form of how to proceed within and beyond the field of HCI.

Participants should have experience in designing or research into profile-based human-computer interfaces.

Participants are asked to write position papers in which they present their opinion related to:

- environments for profiling, like "At home," "At work," and "In public"
- the content, life cycle and definition of user profiles

Position papers should include opinions on:

1. When do profiles have an added value?
 2. What should a profile contain, how much intelligence should it possess?
 3. Implementation issues: multiple profiles, ownership, security etc.
 4. What is the relationship between profiling and HCI?
- Examples of profiles and/or (future) scenarios in which profiles play a crucial role across environments are appreciated.

Position papers can be sent to:

johan.schuurmans@nl.ibm.com