

# Unobtrusive Data Collection for Web-Based Social Navigation

Katja Hofmann ([katja.hofmann@gmail.com](mailto:katja.hofmann@gmail.com)), Catherine Reed ([catherine.reed@csueastbay.edu](mailto:catherine.reed@csueastbay.edu)), Hilary Holz ([hilary.holz@csueastbay.edu](mailto:hilary.holz@csueastbay.edu)) • California State University, East Bay

In initial laboratory studies, *subsymbolic user behavior*<sup>1</sup> has shown promise as a source of information for social navigation. Scalable, unobtrusive methods are needed for acquiring data on subsymbolic user behavior in field studies or live systems. Current methods are not suitable for use outside the laboratory

because they interfere with normal user behavior and environment. We present a method for unobtrusively collecting subsymbolic user behavior in web-based systems, and report results from a field study. Our method is unobtrusive in that it uses current web technologies, works on the vast majority

of current browsers, requires minimal instrumentation of existing web-based systems, and requires no additional user effort. This unobtrusive data collection method paves the way for future research on using subsymbolic user behavior to improve social navigation.

<sup>1</sup> We use the term subsymbolic user behavior to refer to low-level user behavior, that is, users' interactions through mouse and keyboard with a web-based system.

## Subsymbolic User Behavior in Social Navigation

- Promising initial research
  - Time-spent reading (TSR) and time spent scrolling indicate user interest in web pages (Claypool et al., 2001)
  - TSR can improve social navigation in web-based education (Farzan & Brusilovsky, 2005)
  - Subsymbolic features can be used to extract keywords interesting to web page users with higher accuracy than previous approaches (Hijikata, 2004)
- ➔ Subsymbolic user behavior may be a valuable source of information for social navigation
- Problem: How to collect data on subsymbolic user behavior without interfering with user behavior?
- Previously developed methods
  - Require additional user effort (installation of software, changes in browser configuration, restrictions as to which browser to use)
  - Appropriate for laboratory experiments, but not field studies

## Goal

Develop a method for collecting data on subsymbolic user behavior without interfering with normal user behavior



- Developed at CSU EB Adaptive Hypermedia and Assistive Technologies Lab
- Goal: Bridge UNIX experience gap
- Emulates informal education process
  - Contains peer selected, organized and annotated resources
  - Users collaborate directly (comments) and indirectly (list of popular topics and resources)
- Exploratory learning ➔ model users without interfering with normal user behavior

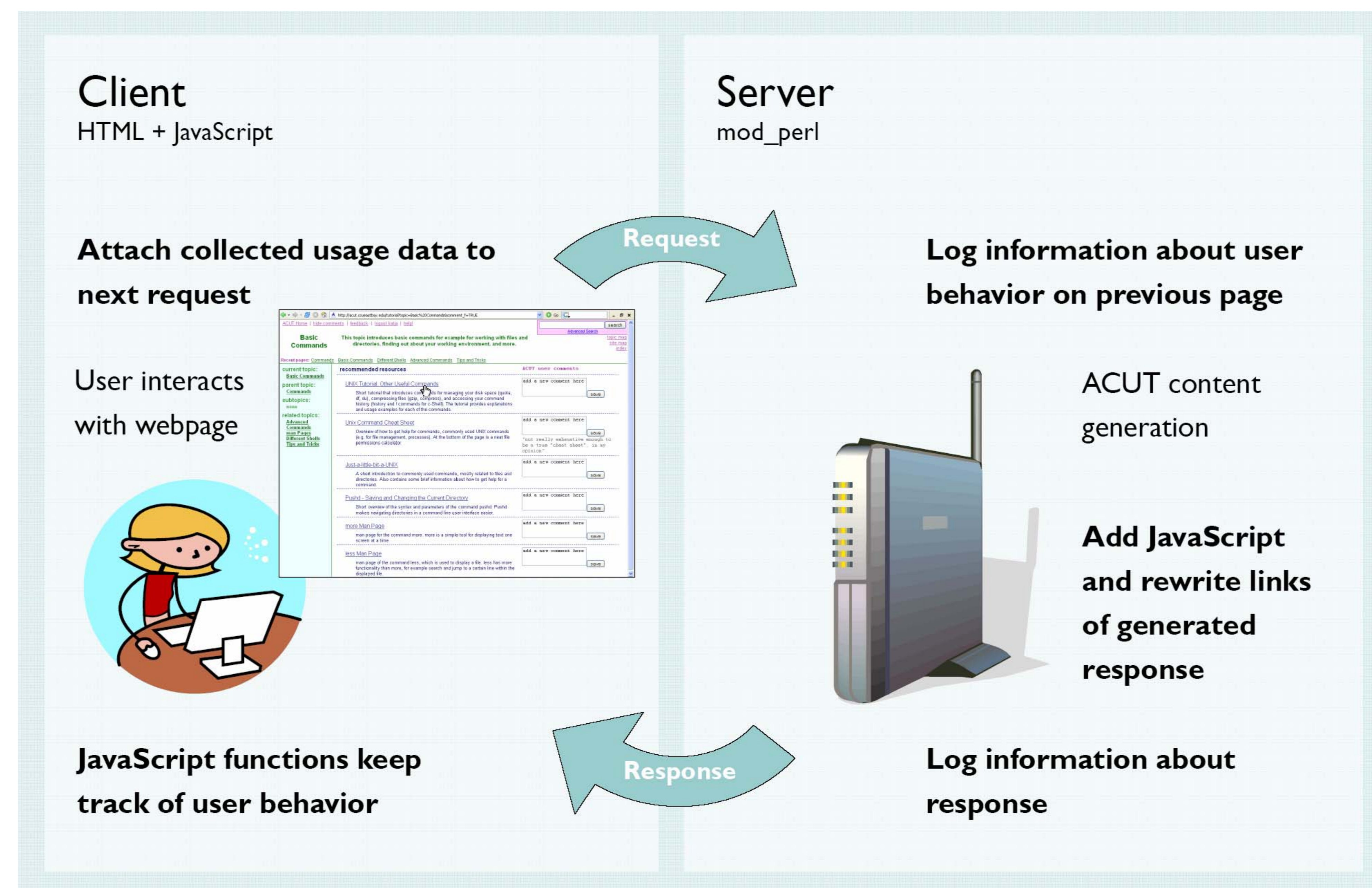


Figure 1: Approach. The Usertrack method for subsymbolic data collection is based on current web-technologies.

## Ongoing and Future Research

- Add support for asynchronous requests to increase coverage
- Collect subsymbolic user behavior data in large scale
- Systematically analyze subsymbolic user behavior data
  - Which aspects of subsymbolic user behavior contain information?
  - What structures or patterns are inherent to subsymbolic user behavior?
  - Can these structures or patterns in subsymbolic user behavior be used for social navigation?
  - Can we identify groups of similar users based on subsymbolic user behavior?
  - Can we infer symbolic information (for example users' goals or effectiveness of interaction with a web page) from subsymbolic user behavior?

## References

- Adaptive Collaborative UNIX Tutorial (ACUT) (2006). Retrieved March 01, 2006 from <http://acc.csueastbay.edu/~acut/>.
- Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. *Proc 6<sup>th</sup> Intl Conf Intelligent User Interfaces (IUI '01)*, Santa Fe, NM, Jan 2001, 33-40.
- Farzan, R. (2003). Adaptive Collaborative Online UNIX Tutorial for Computer Science Students. CSU Hayward, CA. Retrieved February 9, 2004, from <http://acc.csu Hayward.edu/>.
- Farzan, R. & Brusilovsky, P. (2005). Social Navigation Support in E-Learning: What are the Real Footprints? *Proc 3<sup>rd</sup> Wkshop Int Tech for Web Pers*, Edinburgh, UK, Aug 2005, 49-56.
- Goecks, J., & Shavlik, J. (2000). Learning Users' Interests by Unobtrusively Observing Their Normal Behavior. *IUI '00*, New Orleans, LA, Jan 2000, 129-132.
- Hijikata, Y. (2004). Implicit user profiling for on demand relevance feedback. *IUI '04 Funchal*, Madeira, Portugal, Jan 2004, 198-205.
- Kurhila, J., Miettinen, M., Nokelainen, P., & Tirri, H. (2002). Enhancing the Sense of Other Learners in Student-Centred Web-Based Education. *Proc Intl Conf Computers in Education*, Auckland, NZ, Dec 2002, 318-322.
- Svensson, M., Laaksohalmi, J., Höök, K., & Waern, A. (2000). A recipe based on-line food store. *IUI '00*, New Orleans, LA, Jan 2000, 260-263.

## Evaluation

- We employed the Usertrack method for collecting subsymbolic user behavior data in a field study
- Collected data: 30 users; more than 600 page visits; generated more than 5000 events
- Coverage ~ 70% of the pages visited
- Advantages
  - Approach does not interfere with normal user behavior
  - Very high data fidelity on web pages covered (all JavaScript events can be collected)
  - Very general data collection approach; can be added to any web-based system
  - Platform independent: research subjects can use any current JavaScript enabled web browser; server backend can be implemented in any programming language
- Limitations
  - Coverage: Method does not submit collected data to server when user navigates using browser buttons, closes the browser window, or types a new address
  - Pages on which we can collect data need to be retrieved through own web server (but technologies for retrieving third-party content exist)