Computer-Assisted Recording, Pre-Processing, and Analysis of User Interaction Data

Willem-Paul Brinkman
School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom
willem.brinkman@brunel.ac.uk

Philip Gray & Karen Renaud
Department of Computing Science, University of Glasgow, G12 8QQ Scotland
{pdg, karen}@dcs.gla.ac.uk

Although computer-assisted recording, pre-processing and analysis of user interaction behaviour has received continuing research attention over the years, its full potential as a data source to inform the design process seems still unrealised. With technologies as broadband internet and distributed applications, it is possible to continuously and unobtrusively collect interaction data. However, what is the best way to record and collect interaction data? What kind of computer tools, i.e. algorithms, can we use to filter and separate relevant data from noise? And also, which types of analyses and measures give us a design-relevant insight into the interaction, the users, their interaction problems, their needs, personality, and experience? Traditionally, psychologists, usability experts, ergonomists, etc. have been among the main consumers and users of this type of data and supporting tools. However, making the data and tools easy accessible to designers and software engineers might even more directly impact the quality of the application. This workshop plans bringing together individuals with an interest in computer-assisted analysis of interaction data with the aim of improving the interaction design. The workshop will be a platform to discuss new ideas and to share experiences, but also to identify new research challenges and potential solutions. Please note that web usage mining in relation to product sales strategies is not within the scope of this workshop.

Sequential data analysis, logfile analysis, computer-assisted usage analysis, user events.

1. INTRODUCTION

Using the computer to record and analyse user-system interaction is not new. However, with technologies such as broadband in combination with distributed applications, developers now have access to data sets that are: larger in size as they can collect data over a longer time period; more diverse as they can include users from different geological locations; more up to date as the data is almost instantaneously accessible; and more detailed because of increased availability of storage capacity both locally and centrally. The data itself can be used for a variety of things, such as understanding the usability of the system, the systems functionality usage, or the users' needs, personality, and experience. In short, interaction data is and will become an important information resource for developers and researchers. Where other methods only provide subjective results, such as obtained by questionnaires, interviews, and diary studies; or a snap shot, such lab observation or site studies; online tracking of user interaction provides objective data in an unobtrusive manner, which is ideal for longitudinal studies. Despite these clear benefits, it also has a number of unresolved issues. They relate to the way in which interaction data is recorded, collected, pre-processed and finally analysed. This workshop aims to bring together people from the HCI community to discuss these issues and explore potential solutions.

1.1 Recording and collecting usage data
The kind of data and the way it is recorded and collected is not standardised and currently seems very much a matter of individual taste. Some researchers simply record keypresses or cursor movement with a time stamp, whereas others record higher-level or application relevant events [1]. Others have focused on physiological data or other type of behavioural data such as eye gaze. What is recorded seems to relate to the specific implementation of the recording mechanism. For example, if this is implemented in a layer close to the operating system, the recording can include very low-level events [2] such as pressing a single key, or mouse button, or user events that can be understood in the context of the operating system such as creating or killing a process, swapping between applications, opening, resizing, and closing of windows. On the other hand if it is implemented in the application, user events can be recorded in the context of the application, for example a simple press on the enter key becomes a confirmation that a task sequence is completed. A key success factor in the uptake of data collection methods, however, is the ease with which a developer can install this recording mechanism. Several ways seem possible to support developers, such as: providing low-level recording mechanisms tools that are application independent e.g. tracking web access; recording mechanisms that are embedded in the development environment which can simply be activated or deactivated in order to debug a program; or tools for inserting recording routines in specific section of a program [3]. Often tools are still in an experimental phase without any large-scale use, and it is unclear whether any standardisation is possible or desirable. Besides the recording, another important issue is

the collection of the data. Is this first stored locally and transferred centrally in the computers' downtime? Is there local buffering or is the recording directly transported to a central server? The opposite is also possible, with only local recording and no automated collection facilities, and therefore completely depending on the user or the developer to activate the mechanism and submit the data.

1.2 Pre-processing usage data
Once data has been collected, it often needs to be pre-processed. This could mean [4]: 1) selection, separating relevant user events from the 'noise' events; 2) abstraction, relating low-level events with higher-level concepts; and 3) re-coding, generating new event streams based on the results of a selection and abstraction process. Pre-processing seems especially relevant when people are interested in high-level concepts, but only have access to low-level data. Of course this opens the debate about whether we should make high-level tracking easier or whether we should make the tools for pre-processing more powerful.

1.3 Analysing usage data
The final phase is the analysis of the sequential data. Various analysis methods have been proposed [5], such as: Markov analysis, lag sequential analysis, probabilistic finite state machines, Fisher's cycles, maximal repeating patterns, and regular expressions. Whereas these methods are data driven, it is also possible to analyse the data within the context of a model, such as a task model [6], or system-interaction model [7], user models. However, it is unlikely that all potential information about the interaction, the users, their needs, their experiences, their personality etc can already be derived from the data by means of these methods [2]. New analysis methods and supporting tools that can help people to analyse sequential data seems therefore required.

2. WORKSHOP FOCUS

The workshop will focus around the three main topics, which are recording and collecting; pre-processing; and analysis. Both theoretical and practical advances, such as tool development, in these areas can be discussed. The workshop will also consider the users of these tools and methods. Traditionally the target group seems to be highly trained experts such as psychologists, usability experts, ergonomists, etc. However, regarding designers and software engineers as the target user group might increase the impact this data can have on the development of applications. The workshop intends attracting both researchers and practitioners that are interested in sequential data analysis. They could have various backgrounds such HCI, psychology, design, or software engineering. The main objective of the workshop is to establish a community of researchers with an interest in this area, allowing a lively exchange of ideas, and a joint exploration of outstanding problems, but also solutions.

3. WORKSHOP FORMAT

This is a full-day workshop. The morning session will consist of short presentations and discussion of participants' position papers. Participants will also have the opportunity to demonstrate computer-assisted tools they might use or have developed. In the afternoon session, participants will break into small groups depending on their main interest and discuss research questions, unsolved problems, potential solutions and new research directions. At the end of the workshop, the small groups will report back, which will form the basis for a plenary discussion. The workshop will also be supported by a website (http://disc.brunel.ac.uk/HCI2006trackingworkshop) hosted by Brunel University. Position papers will be posted here in advance, and participants' slides will also be made available on the website before the workshop. The intention is also to transform the output of the workshop into a call for a special issue of a journal.

REFERENCES.

[1] Hilbert, D.M. and Redmiles, D.F. (2001) Large-scale collection of usage data to inform design. Proceedings of Interact'01, Tokyo, 9-13 July, pp. 569-576. IOS Press, Amsterdam.
[2] Renaud, K. and Gray, P. (2004) Making sense of low-level usage data to understand user activities. Proceedings of SAICSIT '04, pp. 115-124.
[3] McLeod, I., Evans, H., Gray, P. and Mancy, R. (2005) Instrumenting bytecode for the production of usage data. In Jacob, R., Limbourg, Q. and Vanderdonckt, J. (eds), Computer-aided design of user interfaces IV. (pp. 185-195). Kluwer Acadenic Publishers, Dordrecht.
[4] Hilbert, D.M. and Redmiles, D.F. (2000) Extracting usability information from user interface events. ACM Computing Surveys, **32**, 384-421.
[5] Sanderson, P.M. and Fisher, C. (1997) Exploratory sequential data analysis: qualitative and quantitative handling of continuous observational data. In Salvendy, G. (ed), Handbook of human factors and ergonomics (2ed, pp. 1471-1513). Wiley-Interscience, Chichester.
[6] Lecerof, A. and Paternò, F. (1998) Automatic support for usability evaluation. IEEE Transactions on Software Engineering, **24**, 863-888.
[7] Brinkman, W.-P., Haakman, R. and Bouwhuis, D.G. (2005) Empirical usability testing in a component-based environment: improving test efficiency with component-specific usability measures. Proceedings of EHCI-DSVIS'04, Hamburg, 11-13 July, pp. 20-37. (LNCS 3425) Springer.