

# Paving the Path to Automatic User Task Identification

Anne Gutschmidt  
Graduate School 466\*

Clemens H. Cap  
Computer Science Institute

Friedemann W. Nerdinger  
Business Science Dept.

University of Rostock  
{anne.gutschmidt, clemens.cap, friedemann.nerdinger}@uni-rostock.de

## ABSTRACT

Web site personalization could be immensely improved if the user's current intentions could be recognized by the surfing behavior. The latter can be captured in the form of events occurring in the browser, like mouse moves or opening Web pages. But which aspects of the user's behavior best contribute to the recognition of the task a user is performing? Is it the number of mouse clicks, the amount of time spent on each page, the use of the back button or anything else?

First results of an exploratory study give hint that already simple attributes, such as the average page view duration, the number of page views per minute and the number of different URLs requested, may be usable for the automatic user task identification. 20 participants solved exemplary exercises which corresponded to the user tasks Fact Finding, Information Gathering and Just Browsing. Due to the event logging, true display times were identified, even cached pages and the use of browser tabs were recorded.

## Author Keywords

user tasks, user behavior, interaction, task identification, exploratory study, personalization

## ACM Classification Keywords

J.4 Computer Applications: Social and Behavioral Sciences—*Psychology*; H.5.2 Information Interfaces and Presentations: User Interfaces—*Interaction Styles*

## INTRODUCTION

What if we could recognize the task a Web user is currently performing just by the surfing behavior? The automatic identification of user tasks would improve existing personalization methods by adding a semantic component without explicitly asking the users to give away personal information. Furthermore, users with handicaps, like visually impaired people, would profit from this approach: as soon as the objective is recognized, the user can be supported by a wizard-like program which leads through the next steps to the intended target.

This paper describes a work in progress. It deals with the question which aspects of the behavior are influenced by the task so that eventually an automatic identification of the user task will be possible. This work differs from many existing approaches in the field of interaction tracking, such as [3, 6], in three ways:

1. An exploratory study was conducted in which the user tasks were given in the form of concrete exercises, so that the resulting behavior can be analyzed knowing the real task. Thus, ideal conditions are created: the setting of goals prevents task switching and distractions and makes noise in the data unlikely. It is of particular interest to deduce which task triggers which behavioral pattern.

2. All events are captured which might possibly be of interest, including mouse and scroll moves. Moreover, the selection of browser tabs and the appearance of pop-up windows were considered. Cached pages are also included in the log. This allows a better insight into which Web pages a user really viewed and not only loaded as it is usually the case with Web server logs.

3. It was decided to conduct this study as a pilot study before an extensive field study. That way, a preselection of expedient hypotheses about the relationship between task and surfing behavior can be made under controlled conditions. In comparison with a field study, this pilot study produces a manageable amount of “clean” data allowing a detailed analysis. When starting with a field study right away it may come to the point that the identification of the user task does not work, but the reason is unclear: is it the identification method which does not work or is the input data not applicable or insufficient?

The experiment was conducted using one of Germany's most popular on-line newspapers, *Spiegel Online*.<sup>†</sup> The exercises the participants of the study had to solve represented the user tasks *Fact Finding*, *Information Gathering* and *Just Browsing*, following existing publications like [3].

First investigations showed that there exist differences regarding the page view attributes *average page view duration*, *number of page views per minute*, *number of unique URLs in proportion to the total number of page views* and the *time proportion spent on the start page of the newspaper*.

The total output of the study will be used to formulate hypotheses describing task-dependent behavioral patterns, like “Fact Finding exhibits significantly shorter dwell times than Information Gathering and Just Browsing” etc. Not until enough hypotheses are derived from the study's findings, a field study can be conducted where the surfing behavior of the users will be recorded in a natural surrounding with ev-

\*Supported by the German Research Council (DFG).

<sup>†</sup><http://www.spiegel-online.de>

eryday activities. Thus, the exploratory study described in this paper is an indispensable prerequisite on the way to automatic user task recognition.

The subsequent content is structured as follows: in the next section, user tasks in general and in the context of on-line newspapers are defined based on existing user task taxonomies. Moreover, the state of the art concerning the automatic identification of user tasks is briefly described. The third section is dedicated to the design of the exploratory study. Afterwards, first results of the study are presented followed by a section about related work. The last section summarizes the findings and gives an outlook to future research.

## DEFINITION OF USER TASKS

According to Paternò, tasks are “activities that have to be performed to reach a goal.” A goal is described as “either a desired modification of the state of an application or an attempt to retrieve some information from an application.” Tasks can be divided into subtasks of lower complexity and the relationship between the tasks can be modeled in various ways [7].

In the past, several authors tried to systemize high level user tasks on the Internet [1, 5, 10], but they did not seek a connection to task modeling as suggested by Paternò. Kellar et al. merged these taxonomies of high level tasks with the results of their own study and arrived at the following taxonomy [3]:

*Fact Finding:* The users are looking for a fact in the form of a keyword or a sentence like checking the date of birth of Johann Sebastian Bach; i.e. their target is clearly defined. Such activities are usually of short duration.

*Information Gathering:* The users are collecting information about a certain topic, thus, their target is more open. An exemplary goal might be learning something about baroque music. Due to its research-like character, Information Gathering may take longer, even last for more than one session.

*Just Browsing:* This category describes surfing the Internet with no certain target in mind. It is often of long duration.

*Transaction:* Activities like on-line banking or checking e-mails on-line are summarized in this category.

*Other:* The last category comprises all activities which cannot be assigned to any of the other four categories.

The recognition of these high level tasks would already be a significant breakthrough. Kellar et al. logged the surfing behavior of a group of users who were asked to document their activities during that time period. Events like the usage of the back button, hyperlinks, bookmarks and the history were captured and gathered in log files [3]. One part of the data was used to build a classification rule, the other part was dedicated to testing the rule. The rule was supposed to assign a user to one of the above-mentioned categories (Fact Finding etc.). However, the rule identified only 53.8% of

the activities correctly. The authors claim that this result is caused by individual differences with regard to the surfing behavior [4]. Another reason the authors did not take into consideration is that the differences between Web sites concerning content and structure also have an influence on the surfing behavior.

## THE EXPLORATORY STUDY

### Participants

Twenty students and employees from various institutes of the University of Rostock took part in the test. Their average age was 26.6 years.

### Setting

The participants had to perform exercises on one version of a German on-line newspaper called *Spiegel Online*. The access to external Web pages was blocked. So, the influence different kinds of Web sites can have on the behavior was eliminated and test conditions were as similar as possible for everyone. Each participant underwent the experiment separately, but on the same computer. The Mozilla Firefox browser was used with a software extension for recording all events occurring during the surfing.

### Procedure

In the experiment, the user task was the only parameter which was changed in order to check in which way the behavior changes depending on the task. The participants had to perform exercises which correspond to the user tasks *Fact Finding*, *Information Gathering* and *Just Browsing* following the example of [3]. The category *Transaction* was not adapted as this kind of activity occurs rather infrequently on the newspaper Web site we used. Transactions usually concern article purchases or a newspaper subscription which seemed inappropriate for the test.

At the beginning, the participants were asked to get familiar with the pages of the Web site. They were allowed to surf the Web site as they liked. The maximal duration of this warm-up phase was ten minutes, but the participants were free to decide whether they wanted to finish earlier. This already represented the first exercise and the user task *Just Browsing*.

The two following exercises both corresponded to *Fact Finding*. The first exercise was to look up a certain weather forecast and the second one to find a football result. After having read what they were expected to do, the search was started from the start page of *Spiegel Online*. When the information was found, the participants told the investigator the answer aloud and turned back to the start page for the next exercise.

The last exercise was to collect information on the G8 summit and thus represented *Information Gathering*. The participants were informed that after 10 minutes a few questions pertaining to the topic would have to be answered. This was supposed to act as motivation.

During the experiment, the users' behavior was recorded by capturing every event occurring in the browser. A Mozilla

Firefox extension was developed to log the following events:

- mouse events (clicks, moves and touching page elements)
- scrolling
- keystrokes
- tab events (open, select and close)
- browser events (reload, stop etc.)
- page events (show and hide)

Besides the time of occurrence of the event, further details were saved; e.g. for each mouse click event the information about the element that was clicked was stored; i.e. whether it was a hyperlink and to which page it leads, whether it was a simple text paragraph, a headline, an image or a browser button etc.

At the end, the participants were presented with a *questionnaire* which ascertained demographic information as well as the level of experience concerning computer and Internet usage, the familiarity with on-line newspapers and their browser preferences.

## DATA ANALYSIS

The log files gained from the experiment represent lists of events from which behavioral attributes have to be extracted. Significance tests are done to find differences between the user tasks regarding these attributes. When such a significant difference is found, this attribute may be suitable for the automatic identification of user tasks.

The first investigations concentrated on page views which represent the time a user was looking at a Web page. Page views are derived from the log by considering the page events, i.e. when a page is shown and hidden, but also tab events. Sometimes, users like to load a page in a tab in the background, but this does not actually start a page view as the user cannot yet see the page. The page view only starts when the user selects the according tab.

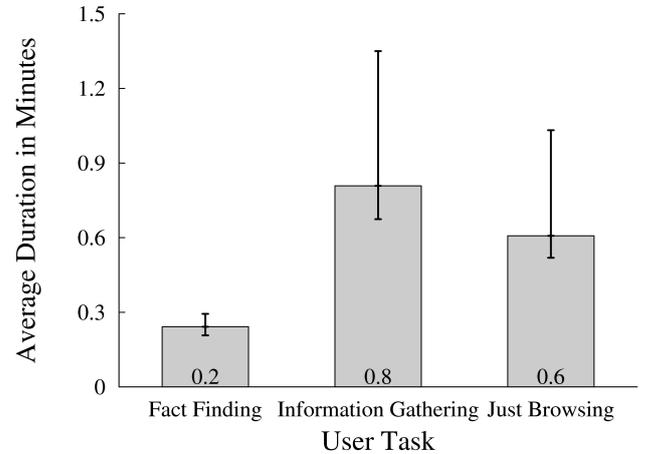
The page views were examined concerning their average duration as well as the number of page views per task and per minute. Furthermore, the time proportion of start page visits during the task and the number of unique URLs (page views without repetition) in proportion to the total number of page views were of interest. A t-test with pairwise samples was done to measure the significance of the difference between the tasks. Usually, t-tests require a normal distribution which cannot be guaranteed here, as the sample size is too small. However, due to the exploratory character of the study this uncertainty is accepted.

## RESULTS

### The Average Duration of a Page View

For each user and each task an average page view duration was calculated. Figure 1 shows the results for the three tasks *Fact Finding* (FF), *Information Gathering* (IG) and *Just Browsing* (JB) including average value and the standard

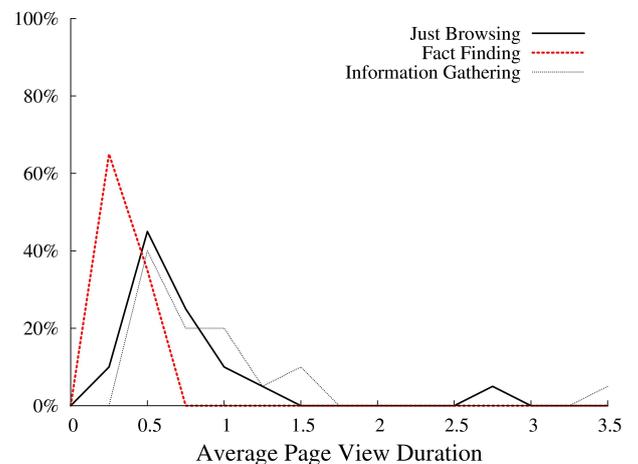
deviation depicted as error bar. The t-test reveals that the difference between each pair of tasks is significant as the significance value is always less than a minimal level of  $p = 0.05$  (see Table 1). The degree of freedom is always  $df = 19$ .



**Figure 1.** A comparison of the average page view duration. ( $\sigma_{FF} = 0.1$ ,  $\sigma_{IG} = 0.7$ ,  $\sigma_{JB} = 0.5$ )

	FF & IG	FF & JB	IG & JB
<b>T</b>	-3.729	-3.143	2.414
<b>p</b>	0.001	0.005	0.026

**Table 1.** Results of the t-test for the average page view duration.



**Figure 2.** The frequency distributions for the average page view duration.

This allows the assumption that Fact Finding takes less time than the other tasks. The high values of the standard deviations suggest that a differentiation between Information Gathering and Just Browsing might turn out difficult. To draw generally valid conclusions and to see if this attribute is reliable for the identification of the tasks, it is necessary to collect data of more participants in a natural surrounding with arbitrary activities. However, these results show that low average duration of a page view is a good clue for Fact

Finding. The investigations will be extended in order to reveal possible patterns; e.g. if there is a development during the session like page views becoming gradually shorter or longer depending on the task.

Figure 2 illustrates the frequency distributions for the three tasks. Their shape allows the presumption of a normal distribution. However, the size of the sample is too small to allow generally valid statements.

### The Number of Page Views

With regard to the number of Web pages viewed during one task, a significant difference is evident between *Fact Finding* and the other two goals. However, a significant difference between *Information Gathering* and *Just Browsing* could not be found as Table 2 shows (see also Figure 3).

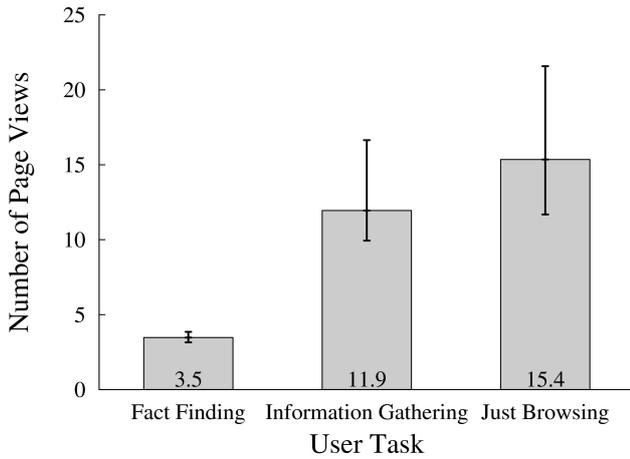


Figure 3. A comparison of the number of page views per task. ( $\sigma_{FF} = 0.7, \sigma_{IG} = 6.7, \sigma_{JB} = 9.9$ )

	FF & IG	FF & JB	IG & JB
<b>T</b>	-5.565	-5.384	-1.400
<b>p</b>	< 0.001	< 0.001	0.178

Table 2. Results of the t-test for the number of page views per task.

	FF & IG	FF & JB	IG & JB
<b>T</b>	7.000	5.293	-1.673
<b>p</b>	< 0.001	< 0.001	0.111

Table 3. Results of the t-test for the number of page views per minute.

A similar result turns out when normalizing this attribute by referring it to a time period of one minute as depicted in Figure 4 and Table 3. This seems quite evident as *Fact Finding* is characterized as a quick search for a well-defined piece of information. Again, this attribute seems to be very suitable for the recognition of *Fact Finding* whereas *Information Gathering* and *Just Browsing* appear similar. Figure 5 indicates a normal distribution for the number of page views per minute.

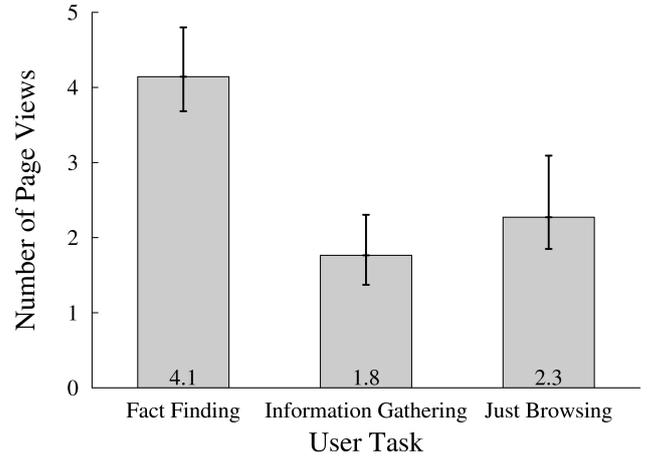


Figure 4. A comparison of the number of page views per minute. ( $\sigma_{FF} = 1.1, \sigma_{IG} = 0.9, \sigma_{JB} = 1.2$ )

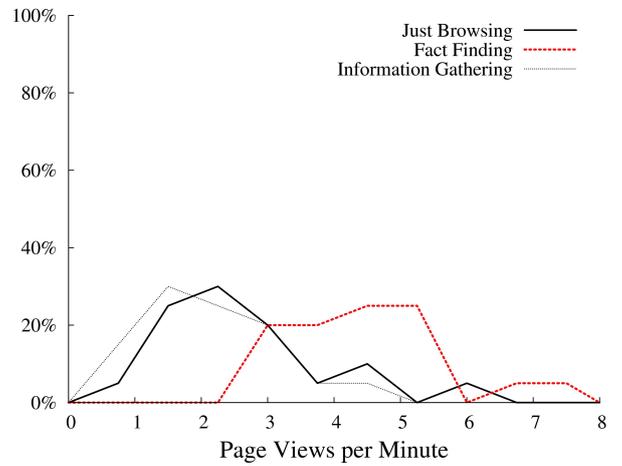
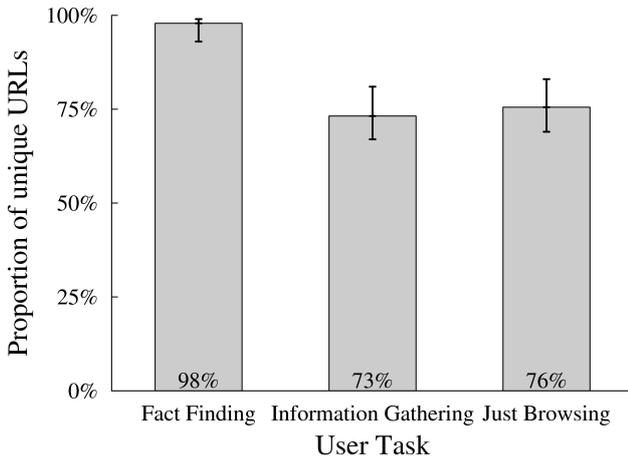


Figure 5. The frequency distributions for the number of page views per minute.

### Unique URLs and Start Page Visits

Two further behavioral aspects were investigated: the number of unique URLs in proportion to the total number of page views and the time proportion of start page visits. The URL proportion reflects whether pages have been visited several times. If the value is high most of the pages have been visited only once. A lower value indicates more repetitions. Again, *Fact Finding* turns out to be significantly different from the other two user tasks with a high average value. In contrast, *Information Gathering* and *Just Browsing* are less easy to differentiate (see Figure 6 and Table 4). The investigations have to be extended in order to find out what causes these repetitions. Maybe one reason is that users navigate deep into the structure and navigate back on the same path to their starting point several times.

The time proportion of start page visits was expected to bring a difference between *Just Browsing* and *Information Gathering*. The t-test confirms this assumption (see Figure 7 and

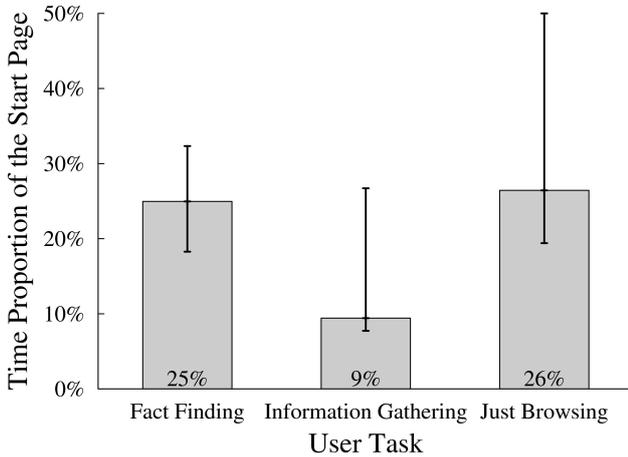


**Figure 6.** A comparison of the number of unique URLs in proportion to the total number of page views. ( $\sigma_{FF} = 5\%$ ,  $\sigma_{IG} = 14\%$ ,  $\sigma_{JB} = 14\%$ )

	FF & IG	FF & JB	IG & JB
<b>T</b>	8.518	7.765	-0.637
<b>p</b>	< 0.001	< 0.001	0.531

**Table 4.** Results of the t-test for the number of unique URLs in proportion to the total number of page views.

Table 5). The start page is the starting point for every task, but for Just Browsing it was expected to be very important as the page contains various headlines from different news categories. Users with different topic interests may spend more time here. In contrast to this, Information Gathering is more specialized. It is probable that users performing Information Gathering use the start page less. This time, it is Information Gathering which is best to differentiate from the other two tasks.



**Figure 7.** A comparison of the time proportion of start page visits. ( $\sigma_{FF} = 14\%$ ,  $\sigma_{IG} = 19\%$ ,  $\sigma_{JB} = 24\%$ )

## Conclusion

With regard to the behavioral attributes described above, differences between the three user tasks occurred. They allow

	FF & IG	FF & JB	IG & JB
<b>T</b>	2.711	0.255	3.051
<b>p</b>	0.014	0.802	0.007

**Table 5.** Results of the t-test for the time proportion of start page visits.

the following hypotheses:

- Fact Finding shows a smaller average page view duration than Just Browsing and Information Gathering.
- Users performing Fact Finding look at more pages during one minute than with Information Gathering and Just Browsing.
- Fact Finders do not tend to page revisits.
- Users performing Information Gathering spend little time on the start page of the newspaper compared to Fact Finding and Just Browsing.

This list is not complete as the similarities between the tasks also have to be included. Furthermore, more detailed investigations have to be done, e.g. on the page view duration as it has already been mentioned.

An issue of concern are the high values of standard deviation for some attributes. Until now, one can only guess that their origin lies in the different motivation of the participants. Some, for example, read very carefully whereas others seemed to be keen on finishing the test. If, however, these differences between the participants are caused by individual surfing habits, this might cause problems with the task identification.

## Limitations

The study collected data of only few users under very controlled conditions. Clearly, this causes some limitations, but these were not accepted without reason. As it was explained at the beginning, we wanted to guarantee that the users are performing exactly the tasks we were expecting in order to have an unambiguous picture of the task-dependent behavior. If the participants were allowed to do what they wanted they could have become distracted or constantly switched between tasks. That way, we gained ideal data in which the behavioral patterns could be searched.

It was decided to conduct the test on only one Web site and exactly one version of it to keep external influences as little as possible. The participants were supposed to see the same and to do the same. However, further tests must include more Web sites. This will show if the differences found so far also occur on other newspaper sites and eventually also on totally different Web sites. We will also include the category *Transaction* as there will certainly be some on-line newspapers requiring this kind of activity.

The list of evaluated attributes is still limited to very simple attributes. More evaluations will be done on the scroll behavior and, connected to this, the information whether a

page was viewed completely or only its beginning. Furthermore, the use of browser buttons like the back button will be analyzed as well as mouse moves; e.g. do Just Browsers click on images rather than on text? This leads to a very important point: until now, the content has not yet been taken into consideration. This information will highly enrich the behavior analysis.

Moreover, the range of evaluation methods will have to be extended. The attributes we have so far derived from the event log will also be used for testing machine learning techniques such as classification trees, clustering or Bayesian models to examine the expressiveness of the attributes.

The subsequent step will be a field study in which the hypotheses gained from the pilot study will be tested under realistic conditions. The event recording tool will be installed on the participants' computers and log their behavior for a period of four weeks. The users will have to document their activities to maintain a connection between the log and the user task. Until now, it was not possible to draw generally valid conclusions due to the small sample size in the pilot study. The field study will bring the amount of data necessary to create and test a method for identifying the user task.

## RELATED WORK

In the second section some state of the art concerning user tasks and user task taxonomies has already been presented. There are, however, further areas besides the Web where the identification of user tasks is of interest.

In the area of workflow management, for example, processes are to be identified which can be compared to user tasks as we have described them here. In [11] it was explained how event logs can be used to create a process model in the form of a Petri net. One event data set shows which kind of workflow the event refers to and the workflow instance as well as the task which is in this context one step in the workflow. An algorithm is proposed which automatically creates process models from such event logs. That way the process structure is revealed enabling its analysis and evaluation as well as appropriate user support through a workflow management system.

[8] also deals with workflows and processes, but focuses on the special character of knowledge intensive work. Knowledge workers cannot profit from the support of workflow management systems as the processes they perform do not fit into rigid process models. Much more flexibility and room for creativity is required. However, similar tasks may reoccur in the future and the information about past workflow instances should thus be maintained and made available for all members of an organization. As the process is not predictable and a fixed model cannot be created, the identification of *process patterns* is suggested. This is done by performing *process mining* using data about the actions that have been performed and the data structures which were involved. This information is obtained by considering the personal knowledge management. The users either have to input the information about their processes on their own,

or the information is unobtrusively gained by observing the user's actions as it is suggested in [9] with the EPOS project. Using different plugins, the behavior within several applications such as e-mail clients, web browsers or text editors is observed to create individual context models. According to [9], these models can be used to infer a user's needs and goals. In contrast to [9] where events from different applications are integrated, the Microsoft Office Assistant refers to user needs within only one application. In Lumière, the Office Assistant's predecessor, every action concerning mouse, scrolling or keyboard is captured [2]. Furthermore the selection of objects within the application, the browsing of menus etc. are observed. A Bayesian model is used to estimate the probability of a user having a certain need.

Most of these approaches do not only deal with the question about what the user is doing, they are, above all, interested in what the user needs, which is closely connected with the user task and the user's goal. However, a final satisfying solution of gaining this information has yet to be found.

## SUMMARY AND OUTLOOK

An exploratory study was conducted to examine the connection between the task a user is currently performing and the resulting behavior. The investigations referred to on-line newspapers to avoid the influence different kinds of Web sites could have on the surfing behavior. Exercises were set that each corresponded to a user task: *Fact Finding*, *Information Gathering* or *Just Browsing*. During the test, every event was recorded in a log file. The first evaluations referred to attributes of page views and already led to promising results. *Fact Finding* was distinguishable from the other two user tasks in four of five attributes, whereas *Information Gathering* and *Just Browsing* showed a difference for only two attributes. The next step will be to examine if the attributes are suitable for the prediction of the task by applying different machine learning techniques such as classification trees or Bayesian models.

The study was to gain hints on which aspects of the behavior might be useful to enable automatic task identification. These hints have to be wrapped in clear hypotheses like "Fact Finders usually view Web pages only briefly, scroll fast, and often do not scroll down to view a page in total." These hypotheses have to be tested in another study which must be more extensive concerning participant number and time. Furthermore, the data should be collected when the users are involved in their everyday activities and not in a laboratory situation.

As soon as a feasible task identification can be implemented, a variety of user support mechanisms can be realized. A recommender system for an on-line newspaper could, for example, use different algorithms for the determination of link suggestions depending on the user task. When dealing with Fact Finding, text mining could be preferred: the user is looking for a small piece of information like a keyword, so we could look for documents containing the same keywords as those Web pages the user has just visited. When dealing with Information Gathering, it would be advisable

to search for recommendations within the same topic frame whereas Just Browsing seems eligible for a method like association mining where articles are suggested that were often read together with those articles the current user has just seen. Depending on the task, different services could also be offered, e.g. a Fact Finder would probably welcome a search functionality provided with the next Web page opened. Furthermore, one could examine whether certain task types regularly occur at certain times of day for a specific user, e.g. in the morning the person often performs Information Gathering, at noon Just Browsing and during the afternoon Fact Finding. Thus, the time of day could already give hint on the task.

The utility of identifying user tasks is, however, not restricted to recommender systems. A live assistant can be developed that recognizes the user task and, connected with this, the user's needs and gives according hints.

The list of utilities that automatic user task recognition could make possible is by far not complete. This underlines even more how important this topic is and will still be in the future.

## REFERENCES

1. Chun Wei Choo, Brian Detlor, and Don Turnbull. Information seeking on the web: An integrated model of browsing and searching. *First Monday*, 5(2), 2000.
2. Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. In *In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265, Madison, WI, 1998. Morgan Kaufmann: San Francisco.
3. M. Kellar, C. Watters, and Shepherd. The impact of task on the usage of web browser navigation mechanisms. In *Proceedings of Graphics Interface (GI 2006)*, pages 235–242, Quebec City, Canada, 2006. Canadian Information Processing Society.
4. Melanie Kellar and Carolyn Watters. Using web browser interactions to predict task. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 843–844, New York, NY, USA, 2006. ACM Press.
5. Julie B. Morrison, Peter Pirolli, and Stuart K. Card. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 163–164, New York, NY, USA, 2001. ACM Press.
6. Laila Paganelli and Fabio Paternò. Intelligent analysis of user interactions with web applications. In *IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces*, pages 111–118, New York, NY, USA, 2002. ACM Press.
7. Fabio Paternò. Task models in interactive software systems. In: Chang, S.K. (Ed.), *Handbook of Software Engineering Knowledge Engineering*. World Scientific Publishing Co, 2001.
8. Uwe V. Riss, Alan Rickayzen, Heiko Maus, and Wil M. P. van der Aalst. Challenges for business process and task management. *Journal of Universal Knowledge Management*, 0(2):77–100, 2005.
9. Sven Schwarz. A context model for personal knowledge management. *Lecture Notes in Computer Science*, 3946/2006:18–33, April 2006.
10. Abigail J. Sellen, Rachel Murphy, and Kate L. Shaw. How knowledge workers use the web. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–234, New York, NY, USA, 2002. ACM Press.
11. W.M.P. van der Aalst, A.J.M.M Weijters, and L.Maruster. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.