Search-Event Recorder (SER) User Guide V1.0

A measurement tool for information behavior in applied contexts

Volker Lang
Marc Schwenzer
Stefan Charonsouk
Search-Event Recorder (SER) User Guide V1.0:
A measurement tool for information behavior in applied contexts

Table of contents

1. Introduction: Purpose of the Search-Event Recorder ......................................................... 3
2. Example: Conducting a study with the Search-Event Recorder ............................................... 4
3. Technical details ....................................................................................................................... 5
   3.1. Structure of the Search-Event Recorder .............................................................................. 5
   3.2. Installing and using the Search-Event Recorder (SER) ...................................................... 6
      3.2.1. Server-side installation ................................................................................................. 6
      3.2.2. Configuration of the browser plug-in ............................................................................ 6
      3.2.3. Client-side installation of the browser plug-in .............................................................. 6
   3.3. Structure of the recorded data and data access ................................................................. 7
4. Conclusion ............................................................................................................................... 8
Bibliography ............................................................................................................................... 9
1. Introduction: Purpose of the Search-Event Recorder

Decisions are grounded in information. Where relevant information is scarce or the quality of given information is low, people tend to step into information seeking processes. By information seeking we mean information search behavior in the sense of Wilson (1999). This kind of behavior consists of two parts: First, strategic search activities, and second, evaluations. Both parts are integrated by documentation activities. In addition, information seeking implies activities clarifying the personal information needs. Together these activities form the active part of human information behavior (see Wilson, 2000).

During the last two decades the use of online information sources has become an important aspect of information seeking (see Case, 2012). This development creates new demands and affordances with respect to the use of related information and communication technologies (ICT, see DiMaggio et al., 2001). Early research on the spread of ICT documented remarkable social inequalities in access and adoption called digital divide (see Katz et al., 2001; van Dijk/Hacker, 2003). Later research found a complementing divide in terms of skills necessary to utilize ICT and preferences of ICT usage (see Hargittai/Hinnant, 2008; van Dijk, 2006; van Deursen/van Dijk, 2011). With respect to the influence of SES and (parental) education on information seeking skills and preferences studies show that these human capital assets are primarily learned outside the classrooms of formal educational institutions (see Becker, 2000; Brandtweiner et al., 2010; Tondeur et al. 2011), especially if these skills are related to digital informational environments (see Attewell/Battle, 1999; van Deursen/van Diepen, 2013). Hence, information seeking skills and preferences tend to be comparatively strongly influenced by informal learning opportunities (see Attewell et al., 2003; Hargittai, 2010), e.g., access to new ICT in the parental household (at early ages; see Prensky, 2001; Wood/Howley 2012). Furthermore, research shows that such opportunities are more frequent if parental SES or education are high (see Chaudhuri et al., 2005; Lang/Hillmert, 2013), making the related informational skills and preferences comparatively strongly dependent on the social background (see DiMaggio et al., 2004). In addition, gender differences are found with respect to the differentiation between actual and perceived ICT skills (more similar for women, see Hargittai/Shafer, 2008) as well as with regard to theoretical ICT competencies (higher for men, but the gap is closing, see Gui/Argentine, 2011).

So far little is known about how these group-specific differences in information seeking skills and preferences influence information seeking processes and related decision behavior. An exception is Hargittai (2002; 2004) who analyzed data on online information seeking behavior acquired by recording the computer screen of study participants. Moreover, it is an open question how information collected for a specific purpose becomes relevant for decisions taken during the further life course. In this paper we describe a data collection tool called Search-Event Recorder (SER) to close the empirical side of this research gap by assessing differences in Internet information seeking processes. The SER records the Internet search processes of users conducted with a standard Internet browser, Google Chrome, which is currently the most frequently used browser worldwide with an estimated market share of 43% (see Statcounter, 2014).
2. Example: Conducting a study with the Search-Event Recorder

SER study participants are invited with an e-mail containing a link to the browser-plug-in and a description of the installation process. After the installation of the plug-in a new button is available in the upper right corner of the Google Chrome browser-window besides the URL-bar. The plug-in is activated by clicking this button. Before the data recording process is started users are asked to login with their username and password. New users have the opportunity to register by clicking on the provided registration link. After a successful login, the SER is active and will send the recorded data to the server. Either loading a new website or changing the focus of the browser-tab will document the previous URL, information on how long the focus of the browser has been on the website and all links contained on the website. An active SER is indicated to the study participants by the red REC-badge on top of the activation button in order to prevent the unintended recording of browsing processes.

To exemplify the procedure assume a study participant who likes to get information on our project and first uses Google to find it by entering the query "wissenschaftscampus tuebingen cluster 4". Afterwards she tries to get information on a specific researcher by clicking on the link to the personal side provided on the homepage of our project. Figure 1 depicts the related search path based on the data recorded by the SER. The nodes are the recorded URLs, the arrows indicate directed links between the browsed websites and the numbers in squared brackets represent the focus duration on the website related to the sender URL of the associated direct link.

**Figure 1:** Example search path based on Search-Event Recorder data
3. Technical details

3.1. Structure of the Search-Event Recorder

The SER consists of the two parts depicted in Figure 2:

1) The client-side JavaScript-based Google Chrome browser plug-in that is capturing the information on the websites accessed.
2) The browser plug-in sends its data to a standard web server that must have the ability to run PHP scripts and contain a MySQL database to store the received data.

To extract the data from the MySQL database an additional MySQL client is needed. Afterwards the data can be analyzed using statistical software packages (e.g., Stata or R).

---

1 We thank Prof. Torsten Grust and his team at the Department of Computer Science at the University of Tuebingen for their support in implementing the SER Google Chrome plug-in.
3.2. Installing and using the Search-Event Recorder (SER)

3.2.1. Server-side installation

In a running server environment the server-side installation is done in three steps.

1. Enter the server configuration:
   In the file "config.inc.php" change the variables so that IP, port, database, user and password specify location and access parameters to the SQL-database.

2. Upload the directory to your webserver (and check afterwards that file permissions are set correctly).
3. Enter the URL path to "install.php" into your browser to initially create the necessary databases. Delete the file "install.php" on success.

3.2.2. Configuration of the browser plug-in

Before the browser plug-in is distributed to the study participants, it needs to be configured to communicate with the server-side PHP-script which demands the specification of the IP address of the server used to store the recorded data.

In order to specify this IP address open the file "pathtoserver.js" in the folder SER of the browser plug-in files and change the text after the variable "pathtoserver" to match the URL-path to the previously installed PHP-scripts and save that file.

Afterwards start up Google Chrome and select the menu "Extension" or type "chrome://extensions/" in the URL-bar and press the "Return"-key. Activate the checkbox "developer mode". Press the button "Pack extension..." and select the path to the folder SER. Finally, specify the path to the folder where the browser plug-in should be written. The generated file ending with "SER.crx" can be sent to the participants of the study for the client-side installation.

3.2.3. Client-side installation of the browser plug-in

The fundamental principal of the SER is that participants have to install the browser-plug-in deliberately in order to prevent the unintended recording of browsing processes.

The client-side installation is done in three steps:
1) After starting up Google Chrome select the menu "Extension" or type "chrome://extensions/" into the URL-bar and press the "Return"-key.
2) The extension which is typically provided through a link is installed via drag and drop into the "Extension" window.
3) The user has to agree to the installation procedure.
After the client-side installation is finished a new button labeled "SER" appears in Google Chrome besides the URL-bar. This button is used to enable or disable the SER.

The browser plug-in can be permanently uninstalled by clicking on the trashcan symbol right of the SER in the "Extension" window.

3.3. Structure of the recorded data and data access.

The SER writes its data into two MySQL tables.

- The table "ser_user" contains the user registration data, consisting of a user identification number called "id", the name of the user and the password used for authorization.
- The table "ser_interaction" contains the recorded meta-data of the browsed websites which are related to specific users by matching the variable "user_id" with the variable "id" in the "ser_user" table. The columns of the "ser_interaction" table consist of the variables shown in Table 1. Whenever browsing a website or switching to another tab a new row is generated in the "ser_interaction" table that contains the data of the previous action.

**Table 1**: Data recorded in the table "ser_interaction" of the Search-Event Recorder

<table>
<thead>
<tr>
<th>Variable</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>ser_id</td>
<td>Identification number randomly generated by the users browser plug-in</td>
</tr>
<tr>
<td>timestamp</td>
<td>Date and time of accessing the website in Unix time format (&quot;%b %d %Y %T&quot;)²</td>
</tr>
<tr>
<td>user_id</td>
<td>User identification number; matches with variable &quot;id&quot; in the &quot;ser_user&quot; table</td>
</tr>
<tr>
<td>url</td>
<td>URL of the website.</td>
</tr>
<tr>
<td>timespent</td>
<td>Duration the browsers was focused on the website (format: &quot;h:m:s&quot;)</td>
</tr>
<tr>
<td>numberoflinks</td>
<td>Number of links contained in the website in progress</td>
</tr>
<tr>
<td>linkList</td>
<td>List of links contained in the website in progress separated by semicolon (&quot;;&quot;)</td>
</tr>
<tr>
<td>useragent</td>
<td>User agent string that specifies the used browser, version and depending on the browser additional information like the used operating system (e.g., &quot;Mozilla/version number (Operating System) Additional Information 1/ Additional Information 2/...&quot;)</td>
</tr>
</tbody>
</table>

² For information on the Unix time format see: http://en.wikipedia.org/wiki/Date_(Unix), last access: 06/17/2014.
The data recorded by the SER is accessible through MySQL clients or the web-client phpMyAdmin. These tools can be used to export the SER data from the MySQL database to comma separated files (.csv-files). E.g., to request an export of the whole "ser_interaction" table one sends the SQL-query "SELECT * FROM ser_interaction" to the MySQL database using a MySQL client. Alternatively it is possible to access the MySQL database from within statistical packages like Stata or R using related MySQL plug-ins.

4. Conclusion

We introduced the SER, a novel data collection tool for studies on online information search processes and related decision behavior. We described the concept of the SER and its technical implementation. In addition, we provided information on how to install the SER and utilize it in applied research. The SER can be used in experimental as well as non-experimental study designs. We hope that the SER will be a useful software tool for research on information seeking behavior and the analysis of decision processes.

---

3 For information on the phpMyAdmin see: http://www.phpmyadmin.net/home_page/index.php, last access: 06/17/2014.
4 For information on using MySQL with Stata see: http://statadaily.ikonomiya.com/2011/05/12/stata-mysql-a-first-encounter/, last access: 06/17/2014.
5 For information on using MySQL with R see: http://cran.r-project.org/web/packages/RMySQL/index.htm, last access: 06/17/2014.
Bibliography

Attewell, Paul; Battle, Juan (1999): Home computers and school performance, INFORMATION SOCIETY 15/1, p.1-10.

Attewell, Paul; Suazo-Garcia, Belkis; Battle, Juan (2003): Computers and young children: Social benefit or social problem?, SOCIAL FORCES 82/1, p.277-296.


Brandtweiner, Roman; Donat, Elisabeth; Kerschbaum, Johann (2010): How to become a sophisticated user: A two-dimensional approach to e-literacy, NEW MEDIA AND SOCIETY 12/5, p.813-833.


Gui, Marco; Argentin, Gianluca (2011): Digital skills of internet natives: Different forms of digital literacy in a random sample of northern Italian high school students, NEW MEDIA AND SOCIETY 13/6, p.963-980.


Hargittai, Eszter (2010): Digital na(t)ives? Variation in Internet skills and uses among members of the "net generation", SOCIOLOGICAL INQUIRY 80/1, p.92-113.


Hargittai, Eszter; Shafer, Steven (2008): Differences in actual and perceived online skills: The role of gender, SOCIAL SCIENCE QUARTERLY 87/2, p.432-447.


Lang, Volker; Hillmert, Steffen (2013): Differential trends in households’ connection to the Internet: An actor-centered explanation, working paper, under review.


Tondeur, Jo; Sinnaeve, Ilse; van Houtte, Mieke; van Braak, Johan (2011): ICT as cultural capital: The relationship between socioeconomic status and the computer-use profile of young people, NEW MEDIA AND SOCIETY 13/1, p.151-168.


