Understanding User Behavior in Digital Libraries Using the MAGUS Session Visualization Tool

Tessel Bogaard[1]([✉]) , Jan Wielemaker[1,2] , Laura Hollink[1] ,
Lynda Hardman[1,3] , and Jacco van Ossenbruggen[1,2]

[1] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
{tessel.bogaard,jan.wielemaker,laura.hollink,laura.hollink,
lynda.hardman,jacco.van.ossenbruggen}@cwi.nl
[2] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3] Universiteit Utrecht, Utrecht, The Netherlands

Abstract. Manual inspection of individual user sessions provides valuable information on how users search within a collection. To support this inspection we present a session visualization tool, Metadata Augmented Graphs for User Sessions (MAGUS), representing sessions in a digital library. We evaluate MAGUS by comparing it with the more widely used table visualization in three representative tasks of increasing complexity performed by 12 professional participants. The perceived workload was a little higher for MAGUS than for the table. However, the answers provided during the tasks using MAGUS were generally more detailed using different types of arguments. These answers focused more on specific search behaviors and the parts of the collection users are interested in, using MAGUS's visualization of the (bibliographic) metadata of clicked documents and selected facets. MAGUS allows professionals to extract more, valuable information on how users search within a collection.

Keywords: Information visualization · Search behavior · Digital libraries · Metadata · Log analysis · User study

1 Introduction

Many studies on large-scale analyses of search logs in digital libraries [2,10,12,19] provide a high-level view of user behavior through methods that report descriptive statistics over groups of sessions, such as demographics, average session duration or number of clicks. Less is known, however, about how search logs can be presented to a researcher or library professional to understand the behavior of individual users. Manual inspection of user sessions (coherent sequences of interactions of an individual user within the search system) provides valuable information on how a user searches within a collection. System developers, for example, inspect sessions to assess whether user behavior on their platform conforms to the system's design. And library professionals are interested in understanding how users search in different parts of the collection to improve search features.

In our research we inspect and interpret user behavior within a historical collection, for instance how users search within different time periods. In the context of a digital library, the documents in the collection are frequently described with rich, professionally curated bibliographic metadata, which can be used to identify users with specific interests [2].

Frequently, a table visualization is used to inspect individual sessions [8]. A table is uncomplicated, typically consisting of a list of queries and URLs of corresponding clicked documents. This, however, has some disadvantages. As an example, in a table it is not directly visible in which part of the collection a user searched; if this is within a specific period, such as World War II (WWII), or for a specific type of document, such as newspaper adverts or family announcements. Also, it can be difficult to recognize specific interaction patterns, such as a user returning to an earlier query, especially in longer sessions.

We present MAGUS (Metadata Augmented Graphs for User Sessions), a tool for visualizing a session in a meaningful way. We describe the design of MAGUS, and discuss in what ways it can overcome the limitations of a table visualization. For example, MAGUS visualizes the facets selected during search and the metadata of clicked documents, providing a visualization of the specific parts of the collection a user is interested in. We evaluate the MAGUS visualization by comparing it with a table representation in three representative tasks completed by 12 participants from diverse professional backgrounds. The questions we address in the evaluation are: (i) *Is session inspection easier in terms of time and effort spent when using MAGUS?*; and (ii) *Are the answers provided better in terms of accuracy and level of detail when using MAGUS?* For transparency, we report all measurements taken, including those that gave negative or inconclusive results, such as agreement between participants or the perceived workload.

2   Related Work

Log Analysis in Digital Libraries. Search logs collected from digital libraries and archives has been studied frequently [2,3,6,10,12,15,16,19]. In some cases, studies focus on the detection and analysis of (topical) user interests, for example to categorize search topics [10,15], or to identify usage patterns in different parts of the collection [2,3]. These studies focus on a statistical analysis of search logs. However, manual inspection of individual sessions can also provide valuable information on how users search in a search system. For example, in [8], individual search behavior is studied to train and develop machine learning algorithms to be able to predict whether a user is demonstrating struggling or exploring search.

Visualization of User Behavior. Frequently used visualisations such as the Behavior Flow in Google Analytics show results aggregated over all users, providing a bird's eye view of search behavior. Similarly aggregated graph visualizations have been used in earlier work, e.g. [4,9]. To visualize a single session, a simple table format is frequently used, e.g. [8]. Alternatively, single sessions have

been represented as linear sequences of colored blocks, with the colors denoting the type of interaction or page visited [13,14,20,21]. In [17], this idea is applied to the search logs of a digital library, with the colors also denoting typical interactions such as adding or removing facets during the search. In this work, we aim to gain more insights into individual user behavior by visualizing single user sessions. We use a directed graph to represent a complete session, and use color and shape of the graph nodes to represent the search and click interactions. The directed graph representation allows the visualization of both the complete navigational path of a user and the repeated user interactions in a single node.

User Studies. In a meta-review of empirical studies focusing on user experience, Pettersson et al., [18], report that in 26% of the studies standardized questionnaires are used, and in 31% user activity is logged, often in combination with other methods, with most studies combining quantitative and qualitative data. In our user study, we similarly combine methods, using activity logging and standardized questionnaires, the NASA-TLX [7] and the System Usability Scale [5], combined with open questions and analysis of answers provided to the tasks.

3   Session Visualization

To visualize a session, we need to specify the start and end of the session, record the queries, facets, and search options submitted during the session and collect information about the documents clicked by the user. For our study, we identify sessions from search logs based on the concept of a clickstream, following the navigational path of a user. The queries, facets, and search options represent the user's search interactions on the platform, and are logged by the search system. Documents in a digital library are frequently described using bibliographic metadata. Clicked documents can be annotated with this metadata, providing insights into the parts of the collection the user searched [3].

3.1   Session as a Table

Table 1. Example table format used by Hassan et al. adapted from [8]

5:55:48 PM	**Query**		Employment issues articles
5:55:52 PM		**-Click**	http://jobseekeradvice.com/category/employment...
6:01:02 PM	**Query**		Professional career advice
6:01:05 PM		**-Click**	http://ezinearticles.com/?Career-Advice-and-Pro...
6:03:09 PM		**-Click**	http://askville.amazon.com/buy-version-Tax-soft...
6:03:35 PM	**Query**		What is a resume
6:04:21 PM		**-Click**	http://en.wikipedia.org/wiki/R%C3%A9sum%C3%A9...
6:07:15 PM			**END OF SESSION**

Sessions are frequently visualized using a table format, typically containing the user queries and URLs of clicked results sequentially, Table 1 and [8]. The format

Date	Query	Info	URL	
"Wed Oct 21 09:11:13 2015"	↳	'click_id=ddd:011108016:mpeg21:a0262'	https://resolver.kb.nl/resolve?urn=ddd:0 ...	1
"Wed Oct 21 09:11:23 2015"	↳	'download_id=ddd:011108016:mpeg21:a0262'	http://www.delpher.nl/nl/pres/view/ocr?i ...	2
"Wed Oct 21 09:11:23 2015"	↳	'download_id=ddd:011108016:mpeg21:a0262'	http://www.delpher.nl/nl/pres/view/cite? ...	3
"Wed Oct 21 09:19:33 2015"	"amersfoort 5 mei 1945"	[]	(-)	4
"Wed Oct 21 09:20:38 2015"	"nsb amersfoort mei 1945"	[]	(-)	5
"Wed Oct 21 09:21:30 2015"	"nsb verzet mei 1945"	[]	(-)	6
"Wed Oct 21 09:21:31 2015"	"nsb verzet mei 1945"	[]	(-)	7
"Wed Oct 21 09:21:55 2015"	↳	'click_id=ddd:010593367:mpeg21:a0288'	https://resolver.kb.nl/resolve?urn=ddd:0 ...	8
"Wed Oct 21 09:50:31 2015"	"nsb verzet mei 1945"	[]	(-)	9
"Wed Oct 21 09:50:46 2015"	↳	'click_id=ddd:010593367:mpeg21:a0288'	https://resolver.kb.nl/resolve?urn=ddd:0 ...	10
"Wed Oct 21 09:51:41 2015"	"nsb 1945"	[]	(-)	11
"Wed Oct 21 09:51:43 2015"	"nsb 1945"	[]	(-)	12
"Wed Oct 21 09:52:49 2015"	"nsb 1945"	"[type=artikel]"	(-)	13
"Wed Oct 21 09:53:16 2015"	"nsb 1945"	"[type=artikel]"	(-)	14
"Wed Oct 21 09:53:34 2015"	"nsb 1945"	"[type=artikel]"	(-)	15
"Wed Oct 21 09:53:44 2015"	"nsb 1945"	"[type=artikel]"	(-)	16
"Wed Oct 21 09:53:59 2015"	↳	'click_id=ddd:010622618:mpeg21:a0311'	https://resolver.kb.nl/resolve?urn=ddd:0 ...	17

Fig. 1. Session from Fig. 2 visualized as a table

is uncomplicated, providing an overview of user queries and clicked results. For our table visualization, we adapt the example for the open web, [8], to the context of a digital library. Our table consists of four columns (see Fig. 1): (i) the timestamps of the interactions; (ii) the user query, or in the case of a click or download, an arrow; (iii) additional information on the search interactions, such as selected facets or search options, or a document identifier for clicks and downloads; and (iv) a link to a clicked or downloaded document.

A table visualization suffers from a number of disadvantages. *Issue 1:* it is difficult to see the connection among interactions other than their time sequence. *Issue 2:* it is not easy to recognize repeated interactions, for example, it is not directly visible when a user returns to an earlier query, for example rows 9 and 10 are equal to rows 7 and 8, Fig. 1. *Issue 3:* it can be hard to view all interactions in a session at once, to see how often each type of interaction occurs, especially for longer sessions. In the context of a digital library, it is difficult to see *issue 4*: which facets users selected during the search; and *issue 5:* the (bibliographic) metadata of the clicked results which can provide meaningful information about the different parts of the collection users are interested in. To address these disadvantages we have developed a session visualization tool, the Metadata Augmented Graphs for User Sessions (MAGUS).

3.2   Introducing MAGUS

In MAGUS[1], a session is visualized as a directed graph where the nodes represent the user interactions, and the arrows the navigational path of the user (addressing *issue 1*). MAGUS is built in the SWISH DataLab environment[1], where Graphviz[2] was used for graph visualization. Figure 2 visualizes a relatively small user session. The session starts at the top, where the gray shape indicates that the user arrived by following a link from an external website, in this case a link from a Facebook post. Through the link, the user arrives directly on a specific article (rectangle). From here, the user performed three interactions, temporally ordered from left to right. The user downloads the OCR text of the

---

[1] Demo and source code available at https://swish.swi-prolog.org/p/magus.swinb.
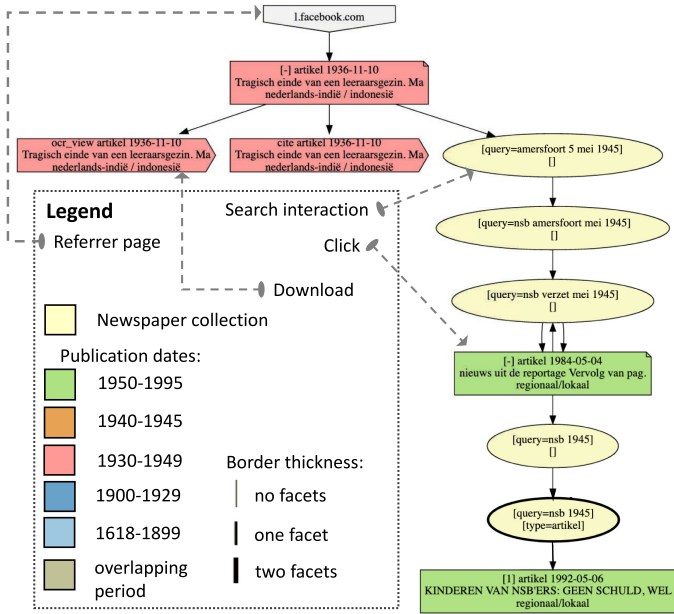[2] http://www.graphviz.org/.

Fig. 2. Session from Fig. 1 visualized with MAGUS (Color figure online)

article, followed by its citation (both indicated by a block arrow shape), then leaves the entry page by initiating a new query and navigating to the search results page (indicated by the yellow ellipse, addressing *issue 3*).

From there, a series of interactions follows: two searches with query refinements (ellipses) and a click on an article (rectangle) are followed by a brief return to the previous page, and back to the article (indicated by the back and forth arrows above the first green rectangle, *issue 2*). To understand the user's search intent it is useful to, in addition to the query, also know which facets were selected (*issue 4*). The user initially used no facets (indicated by the empty square brackets [] in the ellipses), but later added a [type=article] facet, constraining the document type to article (indicated by the thicker line for the last ellipse).

In the historical collection where the example is taken from, it helps library professionals and historians to understand in which period the user is interested. MAGUS allows specific metadata fields to be used to color the nodes in the graph. In this example, we use the publication date from the library's metadata records to color the click nodes. The light red used on the top left indicates documents published in the period around WWII, while the green on the bottom right indicates documents published after 1950 (addressing *issue 5*).

Users exhibit many different interaction patterns, Fig. 3, some of which can be more easily distinguished in MAGUS than in a table. For example, a user clicking from one results page to another using the "next button", or a user selecting
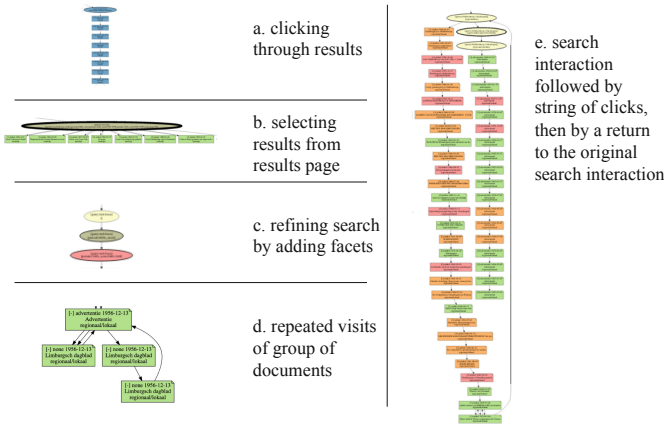
Fig. 3. Multiple graph segments in small size showing different types of user behavior.
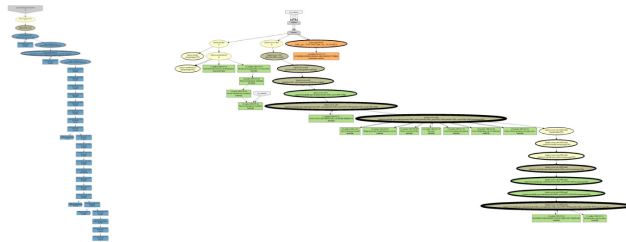

Fig. 4. Two small session graphs. The user on the left was browsing through documents published in the 1900-29 period (succession of blue rectangles). The user on the right was using faceted (thick borders) search interactions (ellipses) after 1950 (green). (Color figure online)

multiple results from the results page and opening them in a new tab, result in deep vertical versus broad horizontal graphs respectively, Fig. 4. Even when the graphs have been reduced in size to a small scale, the difference between the typical "click" behavior of the user on the left can be easily distinguished from the more search-oriented behavior of the user on the right (*issue 3*): the session on the left is dominated by clicks (rectangles) while the session on the right has alternates searches (ellipses) with clicks (*issue 3*). The use of facets is easy to recognize (*issue 4*) in the session on the right by the thick lines used to draw the ellipses of the search nodes, while their color indicates the use of time facets in the post–1950 period (green). The use of the publication dates from the metadata records (*issue 5*) to color the click nodes also immediately conveys that the user on the left is focusing on the 1900–29 period (blue) while the user on the right is more interested in the post–1950 period (green). Additional information about the interaction is displayed in each node. The click and download interactions
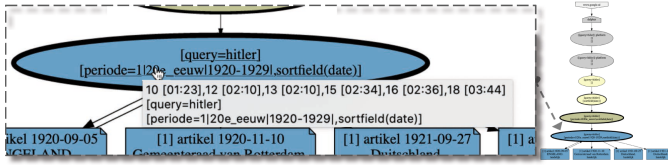
Fig. 5. Hovering over a node displays timestamps with a counter relative to the start of the session, clicking on the node links to the visited web page.

include the document metadata values, the document title, and the page number of the results page of the click. The search interactions include the query, selected facets, and search options used. In addition to the visualizations, hovering over a node will display timestamps relative to the start of the session, and a link to the web page visited (see Fig. 5).

4   Evaluation Setup

In a small-scale experiment we evaluate MAGUS and compare it with a table visualization, Fig. 1. We recruited 12 participants (of which 5 men) among historians, computer scientists, library collection specialists and data scientists. We asked them to perform three tasks and measured the time spent, perceived workload, usability scores (widely used for user studies, [18]). In addition, we measured the certainty of and agreement among the answers given, and performed an analysis of their free-text answers. The experiments were performed on HTTP server logs from the National Library of the Netherlands[3]. The search platform provides access to historical newspaper documents using a faceted interface, with the facets based on the (bibliographic) metadata describing the documents within the collection (such as the publication date). We cleaned and split the logs into sessions as described in [3].

Tasks. The study includes three tasks of increasing complexity. The sessions we selected to be visualized in the tasks all relate to one specific subject–WWII–in the sense that they contain queries and/or clicks on documents about topics related to WWII. This choice is inspired by an ongoing collaboration with the NIOD Institute for War, Holocaust and Genocide Studies[4].

Task 1: identify information needs: *Inspect a session and assess if one of the information needs of the user is to find documents about a topic directly related to WWII.* This task is relevant, for example, to historians who are interested in users searching for WWII-related documents, to understand how users search and which topics they search for. Such a task can also be relevant

---

[3] Logs collected from the search platform https://www.delpher.nl/, access granted under a strict confidentiality agreement.

[4] https://www.niod.nl/en.

to manual label sessions for a training and test set. For example, [11] created such a training set for automatic segmentation of search topics in logs. Each participant performed this task 4 times (subtasks 1.1–1.4).

Task 2 distinguish struggling from exploring users: *Inspect a session and assess whether the user was struggling or exploring.* This kind of task could be performed by a library professional who seeks to understand if users find what they are looking for in the library collection. It is also relevant when building a training set for a classifier, as was done by crowd workers in [8]. Disambiguation between struggling and exploring sessions is important both for understanding search success and when providing real-time user support [8]. Participants performed this task 4 times (subtasks 2.1–2.4).

Task 3 describe a cluster of sessions: *Provide fitting labels and descriptions for four clusters of sessions, by inspecting four sessions per cluster.* In this task, we study to what extent inspection of a few (in this case four) individual sessions allows a professional to see shared, high-level usage patterns and distinguish different types of uses.

For tasks 1 and 2, we manually selected sessions that we judged to be suitable for the tasks and that demonstrate a user interest in WWII topics, based on a list of WWII-related terms provided by the NIOD. For task 3, we clustered sessions including WWII topics using a k-medoids algorithm as described in [2]. This resulted in four distinct clusters. Table 2 provides median values of the clustering features, serving as a high-level overview of the sessions in each cluster. Cluster 1 contains sessions with mainly clicked documents and little search interactions; cluster 2 sessions with clicked documents followed by downloads; cluster 3 sessions with faceted search, focusing on the 1930–49 period; and cluster 4 faceted search with the focus outside the 1930–49 period. In task 3, participants of the study were not shown the session statistics, but were presented with the four most typical sessions of each cluster, i.e. the sessions with the shortest Manhattan distance to the set of medians of the session features in a cluster.

Table 2. Median values of all clustering features for the four clusters.

Cluster	Clicks	Downloads	Search	Search facets	Search WWII facets	Search 1930–49 facets	Search time ranking	Clicks WWII	Clicks 1930–49
1	**88%**	0%	11%	0%	0%	0%	0%	1%	20%
2	33%	**64%**	0%	0%	0%	0%	0%	0%	0%
3	22%	0%	**76%**	**46%**	0%	**11%**	0%	**50%**	**91%**
4	23%	0%	**74%**	**44%**	0%	0%	6%	3%	18%

Two Visualizations. We use a within-subjects design where each participant is exposed to both visualizations. We always present tasks and sessions in the same order. However, we present the visualizations in different orders to avoid measuring a learning effect for either visualization. One group uses MAGUS for subtasks 1.1 and 1.2, the table for subtasks 1.3, 1.4, 2.1 and 2.2, and then

MAGUS for subtasks 2.3, 2.4 and task 3, the other group swaps the visualization tools. Participants are randomly spread over the groups.

Procedure, Data Collection and Data Preparation. First, each participant receives a short training in the use of both visualizations. Then, the participant performs the three tasks. Finally, the participant fills out the System Usability Scale (SUS) questionnaire [5] for both visualizations, and provides further written comments on the use of both visualizations. For the sessions in tasks 1 and 2, the participants select an answer (yes/no on task 1; struggling/exploring on task 2) and provide a free-text justification of their answer. For the clusters in task 3, they provide a free-text label and description. After each session or cluster, we ask participants to assign a measure of their certainty on a five-point Likert scale. After each task and for each visualization method, the participants fill out a NASA TLX questionnaire [7]. All tasks were timed.

We manually annotate the free-text answers to record whether the participants' arguments contain one or more of eight categories of information about a session: (1) queries (for example, a participant writes "hitler as search term"); (2) clicks (for example, "left [...] without clicking"); (3) downloads ("the user didn't download"); (4) links ("possibly saved links"); (5) specific content or metadata values in documents or search facets ("all post-war phenomena" or "time range around ww2 (30–49)"); (6) search behavior ("doesn't use facets", or "click through the results"); (7) blacklist notice, a warning page shown before accessing Nazi-propaganda ("he/she clicked on the blacklist consent"); (8) time ("he/she spent not too much time"). Subjective arguments are left out, such as "he/she seems knowledgeable", "I wonder if they can find it", "couldn't find what he/she was looking for", or "feels more frustrated".

5 Evaluation Results

**Table 3.** Argument analysis of participants' free-text explanations and descriptions.

Task	Visualization	Mean word count	Arg. count	Query	Click	Downl.	Link	Spec. content/ metadata	Search techn./ behavior	Blackl.	Time
1	MAGUS	14	57	25	3	0	0	25	1	3	0
	Table	13	54	37	0	0	0	16	1	0	0
2	MAGUS	26	75	26	16	6	2	6	16	2	1
	Table	25	59	25	12	4	0	2	12	0	4
3	MAGUS	20	55	2	12	7	5	10	19	0	0
	Table	22	58	7	16	10	6	4	15	0	0
Combi	MAGUS	20	**187**	53	31	13	7	**41**	**36**	5	1
	Table	20	171	**69**	28	14	6	22	28	0	4

Free-Text Answers. We analyze the manually annotated free-text answers by counting how many times each argument-category was used by participants. Table 3 shows the number of arguments in total and of each category separately,

for the three tasks as well as overall. It also lists the mean word count of the free-text answers. We notice that only slightly more arguments were used with MAGUS than with the table visualization (187 for MAGUS vs. 171 for the table), and on average the same number of words (20). Only in task 2 participants clearly use more arguments when using MAGUS. However, the type of arguments used is different between the two visualizations. When using the table, participants use the query more frequently as an argument (53 times with MAGUS vs. 69 with the table). With MAGUS, the focus is more strongly on specific content and metadata (41 times with MAGUS vs. 22 with the table), and on search behavior (36 vs. 28). This suggests that MAGUS indeed focuses participants' attention not only on the query but also on other aspects present in the sessions, such as the metadata and the search techniques used.

The free-text cluster descriptions given by participants in task 3, show a difference between MAGUS and the table. As discussed in Sect. 4, cluster 3 focuses on WWII, while cluster 4 does not. Five out of six participants who used MAGUS for task 3 mention this in their description of cluster 3 and/or cluster 4. Only one of the participants that used the table does, labeling cluster 4 as "advanced search after WWII". This demonstrates how MAGUS can improve the quality of answers for tasks where it is important to understand how users search in different parts of the collection.

Agreement Between the Participants. For tasks 1 and 2, we do not consider answers as correct or incorrect, but rather check whether participants agreed on their answers. The number of participants that agreed with each other is exactly the same among participants that used MAGUS and among those that used the table, showing that the visualization method does not impact the agreement. Agreement is different for the different tasks, with almost perfect agreement on task 1 and moderate disagreement on task 2.
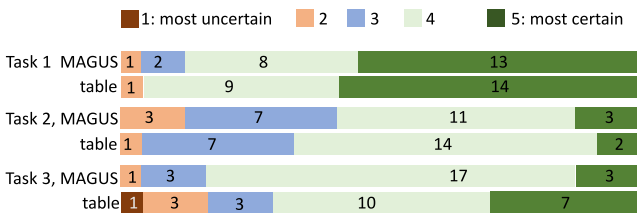
Fig. 6. Certainty: number of times each point on a Likert scale from uncertain to certain was selected.

Certainty of the Answers. We find no differences between MAGUS and the table with respect to how certain participants are of their answers (Fig. 6).
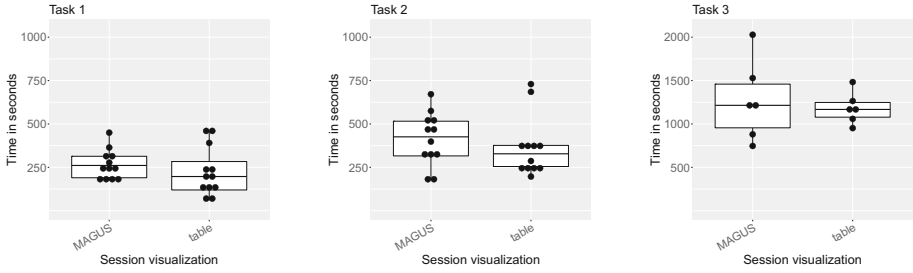
Fig. 7. Time spent per task. Dots represent participants. (Different scale on Task 3.)

Time Spent. The participants need, on average, more time when using MAGUS than when using the table for task 1 and especially task 2. There is no clear difference on task 3. The observed difference in time spent between the two visualizations is small compared with the variation among participants and the difference between tasks, with task 3 requiring considerably more time (Fig. 7).

Workload. Table 4 presents the perceived workload for both session visualizations. Workload is measured through the NASA TLX questionnaire on six dimensions. For task 1, the perceived workload is lower for MAGUS than for the table on all dimensions. For task 2, on the other hand, all workload dimensions are scored slightly higher for MAGUS, and for task 3 the workload is even considerably higher for MAGUS. However, again, standard deviations are high on all questions; variation among participants is generally higher than the difference between the table and MAGUS.

Table 4. Perceived workload measure, on a scale from 0 to 100, lower is better.

| | Task 1 | | | | Task 2 | | | | Task 3 | | | |
| | MAGUS | | Table | | MAGUS | | Table | | MAGUS | | Table | |
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Mental demand	20	37	34	36	51	17	47	22	67	14	52	21
Physical demand	8	9	10	18	11	6	10	5	19	23	13	6
Temporal demand	15	19	21	26	31	18	26	16	41	24	32	23
Performance	20	30	27	28	45	17	40	18	53	12	40	13
Effort	17	32	31	33	43	16	41	18	62	13	34	14
Frustration	14	23	14	18	27	23	23	18	27	18	26	10

Usability. In terms of the reported usability (Fig. 8), the differences are small. MAGUS is liked a bit more than the table. Some participants find the table cumbersome. On the other hand, the participants feel that MAGUS is a bit more difficult to use, as can be seen from the slightly better scores of the table visualization on complexity, ease of use, and the need for support. While the
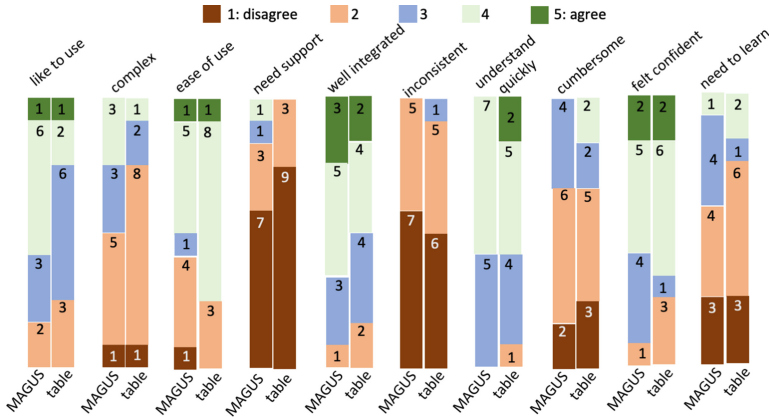
Fig. 8. Usability: number of times each point on a Likert scale from uncertain to certain was selected in the System Usability Scale (SUS) questionnaire.

majority of participants reported that there was little need to learn how to use the two visualizations, multiple participants comment on this. For example, participant 1, an information professional, writes: "You need to learn how to read a graph and understand what is happening in it. But if you inspect it (more) carefully with a legend, then it provides a wealth of information!" We find no conclusive differences with respect to the usability aspects "well integrated", "inconsistant", "understand quickly" and "felt confident."

6   Conclusion

We have developed MAGUS, a tool for visualizing individual user sessions. MAGUS visualizes a user's navigational path as a directed graph, mapping repeated interactions onto a single node. Our tool highlights the different types of user interactions such as searches, clicks and downloads, the use of search facets, and relevant metadata of the clicked documents. In this way, MAGUS allows researchers and library professionals to recognize different interaction patterns and provides insights into the parts of the collection a user is interested in.

We have evaluated our tool on three tasks performed by 12 professionals in a comparison with a standard table visualization. An analysis of the free-text answers demonstrated that MAGUS indeed enabled participants to identify the part of the collection a user is interested in, and that it helps to distinguish different types of search behavior. Further empirical research into specific aspects of the session visualization separately, such as the metadata coloring, could provide more insights into the benefits of each aspect. The results of the workload questionnaire and activity logging suggest that participants find MAGUS more difficult to use than the table, even though the participants do like the tool. MAGUS may be perceived as more difficult due to the steeper learning curve

associated with our tool, and it would be interesting to do a follow-up study to confirm this. A larger follow-up study could also include an investigation of the different professional backgrounds of participants, for example, to compare whether data professionals and domain experts use the tool differently. Furthermore, we would like to investigate which types of tasks specifically benefit from MAGUS, and for which types of sessions the tool works best, as several participants mentioned the benefit of MAGUS for long, complicated sessions.

References

1. Bogaard, T., Wielemaker, J., Hollink, L., van Ossenbruggen, J.: SWISH DataLab: a web interface for data exploration and analysis. In: Bosse, T., Bredeweg, B. (eds.) BNAIC 2016. CCIS, vol. 765, pp. 181–187. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67468-1_13

2. Bogaard, T., Hollink, L., Wielemaker, J., Hardman, L., van Ossenbruggen, J.: Searching for old news: user interests and behavior within a national collection. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval CHIIR 2019, pp. 113–121. ACM, New York (2019). https://doi.org/10.1145/3295750.3298925

3. Bogaard, T., Hollink, L., Wielemaker, J., van Ossenbruggen, J., Hardman, L.: Metadata categorization for identifying search patterns in a digital library. J. Doc. **75**(2), 270–286 (2019). https://doi.org/10.1108/JD-06-2018-0087

4. Brainerd, J., Becker, B.: Case study: e-commerce clickstream visualization. In: Andrews, K., Roth, S.F., Wong, P.C. (eds.) IEEE Symposium on Information Visualization 2001 (INFOVIS'01), SanDiego, CA, USA, 22–23 October 2001, pp. 153–156. IEEE Computer Society (2001). https://doi.org/10.1109/INFVIS.2001.963293

5. Brooke, J., et al.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)

6. Clough, P., Hill, T., Paramita, M.L., Goodale, P.: Europeana: what users search for and why. In: Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., Karydis, I. (eds.) TPDL 2017. LNCS, vol. 10450, pp. 207–219. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67008-9_17

7. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology, vol. 52, pp. 139–183. Elsevier (1988)

8. Hassan, A., White, R.W., Dumais, S.T., Wang, Y.M.: Struggling or exploring? Disambiguating long search sessions. In: Seventh ACM WSDM (2014). https://doi.org/10.1145/2556195.2556221

9. Hong, J.I., Landay, J.A.: WebQuilt: a framework for capturing and visualizing the web experience. In: Proceedings of the 10th International Conference on World Wide Web WWW 2001, pp. 717–724. ACM, New York (2001). https://doi.org/10.1145/371920.372188

10. Huurnink, B., Hollink, L., Van Heuvel, W.D., De Rijke, M.: Search behavior of media professionals at an audiovisual archive: a transaction log analysis. J. Am. Soc. Inform. Sci. Technol. (2010). https://doi.org/10.1002/asi.21327

11. Jones, R., Klinkner, K.L.: Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management CIKM 2008, pp. 699–708. ACM, New York (2008). https://doi.org/10.1145/1458082.1458176

12. Jones, S., Cunningham, S.J., Mcnab, R., Boddie, S.: A transaction log analysis of a digital library. Int. J. Digit. Libr. **3**(2), 152–169 (2000). https://doi.org/10.1007/s007999900022

13. Lam, H., Russell, D.M., Tang, D., Munzner, T.: Session viewer: visual exploratory analysis of web session logs. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2007, Sacramento, California, USA, 30 October–1 November 2007, pp. 147–154. IEEE Computer Society (2007). https://doi.org/10.1109/VAST.2007.4389008

14. Liu, Z., Wang, Y., Dontcheva, M., Hoffman, M., Walker, S., Wilson, A.: Patterns and sequences: interactive exploration of clickstreams to understand common visitor paths. IEEE Trans. Vis. Comput. Graph. **23**(1), 321–330 (2017). https://doi.org/10.1109/TVCG.2016.2598797

15. Meij, E., Bron, M., Hollink, L., Huurnink, B., de Rijke, M.: Mapping queries to the linking open data cloud: a case study using DBpedia. J. Web Semant. **9**(4), 418–433 (2011). https://doi.org/10.1016/j.websem.2011.04.001

16. Niu, X., Hemminger, B.: Analyzing the interaction patterns in a faceted search interface. J. Assoc. Inf. Sci. Technol. (2015). https://doi.org/10.1002/asi.23227

17. Niu, X., Hemminger, B.M.: A method for visualizing transaction logs of a faceted OPAC. Code4Lib J. **66**(5), 1030–1047 (2010)

18. Pettersson, I., Lachner, F., Frison, A.K., Riener, A., Butz, A.: A Bermuda Triangle?: a review of method application and triangulation in user experience evaluation. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI 2018, pp. 461:1–461:16. ACM, New York (2018). https://doi.org/10.1145/3173574.3174035

19. Walsh, D., Clough, P., Hall, M.M., Hopfgartner, F., Foster, J., Kontonatsios, G.: Analysis of transaction logs from national museums Liverpool. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) TPDL 2019. LNCS, vol. 11799, pp. 84–98. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30760-8_7

20. Wei, J., Shen, Z., Sundaresan, N., Ma, K.L.: Visual cluster exploration of web clickstream data. In: 2012 IEEE Conference on Visual Analytics Science and Technology, VAST 2012, Seattle, WA, USA, 14–19 October 2012, pp. 3–12. IEEE Computer Society (2012). https://doi.org/10.1109/VAST.2012.6400494

21. Zhao, J., Liu, Z., Dontcheva, M., Hertzmann, A., Wilson, A.: MatrixWave: visual comparison of event sequence data. In: Begole, B., Kim, J., Inkpen, K., Woo, W. (eds.) Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, Seoul, Republic of Korea, 18–23 April 2015, pp. 259–268. ACM (2015). https://doi.org/10.1145/2702123.2702419