

Interacting with Digital Documents

A Real Life Study of Historians' Task Processes, Actions and Goals

Laura Korkeamäki[†]

Faculty of Information Technology and Communication
Sciences, Tampere University
Tampere, Finland
laura.korkeamaki@tuni.fi

Sanna Kumpulainen

Faculty of Information Technology and Communication
Sciences, Tampere University
Tampere, Finland
sanna.kumpulainen@tuni.fi

ABSTRACT

The aim of this study was to find out what is historians' task-based information interaction like in digital environments. The study was based on TBII framework [20]. The research was conducted in real-life setting using interview and shadowing. The following task process types were identified: (i) searching secondary sources, (ii) reading and making notes, (iii) collecting and processing data, (iv) analyzing and (v) writing. We found that IR interfaces were not always optimal for historians' use, and that the interfaces should support the expression of information needs better. The data processing consumes a lot of historians' time and they could benefit from designing better tools. We also found out that task process types have an effect on how digital tools or data are being used. Therefore, the context in which the tool or data is being used, should be taken into account when designing tools for historians.

CCS CONCEPTS

- Information systems~Users and interactive retrieval

KEYWORDS

Information interaction, Task-based studies, User studies, Ethnographic methods

ACM Reference format:

Laura Korkeamäki and Sanna Kumpulainen. 2019. Interacting with Digital Documents: A Real Life Study of Historians' Task Processes, Actions and Goals. In *Proceedings of ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR' 19)*. 10-14 March, 2019, Glasgow, Scotland, UK. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3295750.3298931>

[†]Corresponding author

1 INTRODUCTION

People interact with information in work and leisure contexts, and the goal is often to use that information in real-life tasks [4]. Much of the information interaction takes place in a digital environment because technology is becoming ubiquitous in people's lives. A digital environment can be defined as "a context, or a 'place', that is enabled by technology and digital devices, often transmitted over the internet, or other digital means, e.g. mobile phone network" [33]. In this research, we aim at studying historians' information interaction during work task performance in digital environments. We focus on historians' task processes and, more precisely, on historians' task goals and actions.

As early as 1990, Bates argued that people want more control over search process, and information retrieval (IR) systems should support the search strategies of people's own choosing – not the other way around [2]. Although digital technology and IR systems have developed vastly since then, we are still dealing with the same issues. Especially for historians, it is vital to build contextual knowledge around the research questions [12]. Therefore, it is not enough for historians to form search queries and trust the algorithms to do their magic. Instead, historians have a habit of combining several search strategies [1, 9, 11, 31].

Information interaction happens in all parts of the research process, not just during information searching. In this research, we use task-based information interaction (TBII) framework presented by Järvelin et al. [20] including activities from task planning to synthesizing and reporting. Studying information interaction is important because digital environments and tools can change how people interact with information [4]. For example, when new digital tools are created for historians, the designs can have an effect on how data can be viewed and what kind of connections historians can see in the data [14]. Consequently, digital tools affect what kind of history research can be made.

In this research, we focus on historians' goals and actions during work task performance. The goals within the tasks can change and alter the task process, even when the task itself remains the same [34]. Therefore, it is important to study task goals and related actions in order to support researchers' work.

2 BACKGROUND

2.1 History as a Domain

History research helps to understand the present based on previous experiences. In information studies, information seeking of humanities and history scholars have been a long-lasting interest [3, 13]. As an academic domain, history is sometimes connected to humanities and sometimes to social sciences [5]. There are many subfields in history, e.g., political, social and culture history.

Historians study the past by exploring various kinds of primary sources, such as letters, diaries, newspapers, magazines and official reports. Works published by other researchers (e.g. books, essays, articles) are considered secondary sources [7]. Searching for primary materials may require a trip to an archive, although historians are increasingly relying on digital primary sources [9, 30]. The varied information environment, the use of primary sources and the emerging digital methods makes history a very timely and interesting research domain in information interaction.

According to Brundage [7], history should not be viewed as a static concept, because “historians constantly search for fresh sources, approaches, methodological tools, and interpretations, in an effort to offer an ever-new past to whatever the present is”. In addition, Cullen [10] pointed out that the ability to form good questions, while writing an essay or doing research is particularly important in history. These ways of working require interacting with information, which is the focus of our research.

There are several studies concerning historians’ search for primary sources [1, 9, 31]. Dalton and Charnigo [11] covered both primary and secondary sources. Duff and Johnson [12] studied historians’ information seeking behavior. In addition, Martin and Quan-Haase [24-26] have written several research papers of historians’ information behavior and serendipity in historical research. There are also studies of historians’ [30] and humanists’ [18] research practices in digital environments. However, information interaction research in the domain of history is lacking. Although Freund and Toms [16] studied how genealogists and historians interacted with archival finding aids, their research was conducted in a laboratory-like setting. Here we address historians’ information interaction in real world settings.

2.2 Conceptual Framework

In this research, we use the following task-related concepts: *activity*, *task process*, *action* and *task goal*.

Firstly, by using the term *activity* we refer specifically to the activities presented in *TBII framework* [20]. TBII framework focuses on cognitive and behavioral aspects in human performance divided into five generic activities (Figure 1). The first activity is *task planning and reflective assessment*, which is viewed as a meta-activity, ongoing through the whole task process. It includes e.g., setting goals, making plans and altering them if necessary.

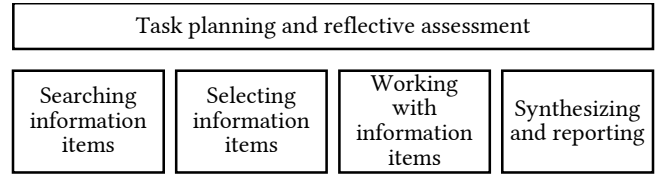


Figure 1: Five activities in TBII framework [20]

Searching information items is a relatively brief and possibly a repeated activity. The purpose of this activity is to find information items to be used in the task process. The term *information item* refers to books, articles, paragraphs, pictures, videos etc. In the activity of *selecting information items* focus is put on deciding which information items are relevant in performing the task. *Working with information items* includes e.g., reading the information items, organizing them, making analysis or converting them into more suitable formats. *Synthesizing and reporting* involves e.g., writing research articles, making conclusions and presenting the results to others.

Secondly, according to Byström and Hansen [8], a *task process* means different kinds of *actions* that an individual performs over time in order to reach a goal. These actions can be cognitive, behavioral or affective [21].

Thirdly, the importance of understanding people’s *task goals* and activities has been widely recognized. For example, Blandford and Attfield [4] noted, “information is only useful when it is interpreted and applied by people in the context of their goals and activities”. With the help of Daniels’ classification of goals, Xie [34] proposed four levels of user goal types: long-term goal, leading search goal, current search goal and interactive intention. It is important to be aware of that task goals can differ depending on how general or specific they are. In our research, the term *goal* refers to participants’ perceived task goals.

Lastly, the concept *action* can be described at different levels. For example, Bates [2] introduced the following concepts related to search tasks: moves, tactics, stratagems and strategies. However, Bates’ definition of the term “move” would be too specific for the purpose of this research. Hence, we use the term “action” in a broader sense. A set of one or several actions form a task process.

3 RESEARCH DESIGN

3.1 Research Questions

We aim at studying historians’ task-based information interaction in digital environments by utilizing the TBII framework [20]. Digital environment was seen relevant, because the use of various types of digital sources have gained increasing attention among historians.

We study the activities from TBII model as task performance processes. The research questions are derived from the view presented by Byström and Hansen [8], who suggested that task processes consist of certain goals and actions. Historians’ task-

based information interaction in digital information environments is studied by addressing the following research questions:

RQ1. What are *the task process types* of historians?

RQ2. What kinds of *goals* are included in the different task process types?

RQ3. What kinds of *actions* are included in the different task process types?

3.2 Methods

We used a qualitative research setting that is discussed next. This is a naturalistic field study. The data collection methods were interview and shadowing, which were conducted in the premises of three universities in Finland. Data was collected between January and March 2018.

3.2.1 Recruitment and participants

The participants were selected by purposive sampling. The criteria for sampling were the following: (i) participants should use digital primary sources in their research projects, (ii) they should include experienced and less experienced researchers, (iii) they should include representatives from different subfields in history, and (iv) they should be located geographically close enough because conducting the study required travelling from the researcher.

At first, participants for this study were sought by checking ongoing research projects in digital history from Finnish universities' web pages. Participants were first contacted by email. During recruitment, also snowball sampling was used, meaning that historians were asked to recommend suitable participants for this study. Participants were recruited also from a digital history seminar (organized early 2018), where it was possible to meet historians in person.

This study included five historians (Table 1) who met the criteria of purposive sampling. The participants were doctoral students or postdoctoral researchers. They had worked as researchers approximately from 1 to 12 years, which indicates various levels of expertise. The participants' research topics were related to one or several subfields, i.e., conceptual or cultural history, digital humanism, history of authorship, intellectual, labor or social history.

Table 1: Participants' research experience

Participant	Career stage	Years working as a researcher*
P1	PhD student	1
P2	PhD student	3
P3	Postdoc	12
P4	Postdoc	7
P5	Postdoc	7

*after master's degree

3.2.2 Interview

Written consent was obtained for the interview and shadowing. Semi-structured interviews were conducted in person. Each of the participants was interviewed once before shadowing (Table

2). Interviews lasted 49 min–1.5 h each. By meeting the participants before shadowing the researcher gained better understanding of the participants' research topics and information environment. Interviews were audio recorded and later transcribed.

Table 2: Interviews and shadowing sessions

Participant	No. of interviews	No. of shadowing sessions	Avg. duration of a shadowing session (min)
P1	1	4	140
P2	1	4	73
P3	1	0	0
P4	1	3	151
P5	1	4	127

3.2.3 Shadowing

Shadowing is a research method, where shadowees are observed while doing their real-life work tasks. Shadower can also ask questions from the shadowees during the sessions if needed [27]. Shadowing reveals information that shadowees forgot to mention during interviews or didn't consider worth mentioning. However, it should be noted that the data collected by shadowing is selective regarding time and place. [28] Four of the participants were shadowed three to four times. Each shadowing session lasted about one to three hours. The shadowing sessions included usually one or two task sessions. Data was collected by handwritten field notes and by video recordings of conversations and participants' computer screens during shadowing sessions.

3.2.4 Data processing and analysis

Table 3 shows an overview of the data collected. The analysis was qualitative because in that way it was possible to gain a holistic view on the historians' task processes, goals and actions. First, the historians' different task process types were identified from the shadowing data (cf. RQ1). In this research, different task processes were assigned to different types according to their qualitative features. In other words, a *task process type* refers to a group of similar sequences of performed activities that are clearly observable or can be expressed in words by the participants, when asked to describe their task performance. The task process types were formed pragmatically. For example, one historian was reading a scientific article and making notes for future use. These types of task processes were labeled as "reading and making notes".

Table 3: Material collected and processed for this research

5 interviews	5 h 40 min of audio recordings and 55 pages of word by word transcriptions
15 shadowing sessions	49 pages of transcribed field notes, 30 h 10 min of video recordings and 48 pages of transcribed video material*

*Only relevant parts of the videos were transcribed

Secondly, the goals and actions of each task process type were analyzed (cf. RQ2 and RQ3). In the beginning of the shadowing sessions, the historians were asked to tell the goals of their current tasks and how they were going to reach those goals. The identifications of the goals and actions were supported by observations, additional questions asked during the shadowings, and by the spontaneous conversations with the historians. For example, one historian’s goal for reading an article was to increase basic understanding of a topic. This goal was labeled as “learning”. During the shadowing session, the historian read the article linearly (labeled as “reading”), saved article’s reference information to a reference management program (labeled as “organizing notes and citations”), and wrote notes on paper and to the reference management program (labeled as “making notes”). In addition, the historian noticed that one in-text citation was mentioned several times in the article, and decided to look that source more closely (labeled as “following citations”). The historian explained that getting to know the conversations is helpful when the topic is unfamiliar (labeled as “learning”).

The analysis of the shadowing data was complemented by the data from the interviews. The interviews were analyzed on the parts that were related to the task process types identified from the shadowing data.

We focused on three TBII activities: searching information items, working with information items, and synthesizing and reporting. In this research, selecting information items was considered a part of searching process and was not analyzed separately. Järvelin et al. [20] noted also that selecting information items is interleaved with the searching activity. Therefore, we combined these TBII activities into one category – searching and selecting information items.

Planning and reflective assessment was not the focus in this research because there were relatively few and brief actions observed (e.g., checking to do -lists, keeping a research diary, thinking of a time-effective solution to a task) related to this activity.

4 RESULTS

All historians used digital documents (digital primary and secondary sources) in their work. Some of the historians also used digital research methods. Because many of the historians were currently working with multiple research projects the shadowing data included work tasks from different projects and project stages. Overall, five task process types were identified from the shadowing data (Table 4).

The task process types were partially overlapping. For example, historians often worked with primary sources throughout the research process.

[...] This is that sort of history research that doesn’t divide clearly into those kinds of stages where you first search the primary sources, then you get to know them, sit down and start writing. Instead, you work with the primary sources all the way. (P1, interview)

Table 4: Task process types according to activities in TBII framework

Searching and selecting information items	Working with information items	Synthesizing and reporting
(i) Searching secondary sources	(ii) Reading and making notes (iii) Collecting and processing data (iv) Analyzing	(v) Writing

The search and selection of primary sources were not studied in this research because historians had already collected them for their studies or they knew where to refine them if needed. Therefore, the first (i) task process type in Table 4 concerned only secondary sources, such as books and articles. The task process types three (iii) and four (iv) concerned mainly primary sources. Reading (ii) and writing (v) included interaction with both primary and secondary sources.

4.1 Searching and Selecting Information Items

Historians searched for secondary sources (e.g. scientific articles) using information retrieval systems (e.g. Ebsco, JSTOR, Google Scholar and Finna, which is a Finnish search portal). The goal for searching secondary sources was to support the writing process. Sometimes, at the early stages of the writing process, the information need was vague. Historian P4 said that while searching new information the goal was also finding better focus for the article.

[...] it could be beneficial to this work [writing an article] to do kind of wider searches [...] that in a certain way you also try to find focus for the work. (P4, shadowing)

Historian P5 did complementary literature searches for the current writing task, but was not satisfied with the search results. P5 hoped for more options for expressing ones’ information needs by, e.g., narrowing down the search to the domain of history or to a certain historical era.

It feels like that if the [historical era] is in the headline of a book or an article, then one might be able to find [relevant books or articles] with a bit of luck. [...] one old [library search portal] had it like as an index term, like history in general or a certain [century]. It would help finding. (P5, shadowing)

In general, historians felt that following leads was a good way to find secondary sources. For historian P3 following citations was maybe even more important search strategy than using information retrieval systems.

I do not remember the last time I searched for some library database [...] at least compared to that I just read what other people are referring to. [...] Then the question is how the newest literature comes to my attention. Maybe it is more like those email lists and Facebook [...]. One just sort of intuitively picks what is relevant. (P3, interview)

Historians found secondary sources also in conversations with other researchers and by attending conferences and seminars. In addition, historians browsed sometimes books in libraries. Historian P4 mentioned keeping an eye on secondary sources also for future needs and for personal collections.

During shadowing sessions, selecting information items happened in iterative circles. As participants searched secondary sources, their understanding of the literature improved which helped them identify relevant secondary sources. For example, while using Google Scholar, historian P4 noticed a familiar book and that the author might have other interesting works. Therefore, P4 continued using the author's name as search terms.

I realized that [...] That [the author] is such that appears in several places and who has written interesting stuff. So I just realized that, aa, I should also take these into considerations. (P4, shadowing)

It remained unclear how many of the items the participants selected during literature searches would end up in their final work. For example, at the end of one shadowing session, P5 assessed some of the items potentially useful.

Maybe I did find a few, but then they didn't contain very useful things I needed here. At least to that one footnote I didn't get anything more. But otherwise maybe some details that I can add there and compare. (P5, shadowing)

4.2 Working with Information Items

4.2.1 Reading and making notes

At the beginning of a shadowing session, historian P1 stated that the goal for reading an article was to increase basic understanding on a particular topic. For the same reason, P1 also followed citations during reading. Historians followed citations during reading also because in that way they were able to find more secondary sources.

[the topic of the article] is a bit unfamiliar to me, so I have to check a bit what are the conversations that they are referring to. (P1, shadowing)

Then of course, if I find some new citations [while reading books], perhaps I will also google and save them for myself. (P5, shadowing)

Historians read both electronic and printed secondary sources. For example, one historian preferred reading electronic sources with reader software, because it enabled to write notes directly into the software and save them later in Word format. Some historians preferred printing interesting articles.

Historians made notes for writing tasks. One historian used footnotes in the note making in such a way that they could be easily moved into drafts. Another historian chose to write notes directly into the draft presentation. Sometimes notes were made with regard to future use. For example, historian P5 anticipated future needs.

Then I'll have these notes for the rest of my life. Then I won't have to go back reading these books. And then I can use them while I write my own stuff. (P5, shadowing)

Historians used different kinds of tools in note making, such as note-taking software, reference management software or word-processing programs. Some historians chose also to use cloud-based file storage services (e.g. Google Drive, Evernote, Dropbox) because then the notes could be accessed from any computer. Some also felt that electronic notes were easier to store and refind. On the other hand, the traditional note making method with paper and pencil was still in use. For example, one historian wrote notes in a physical notebook, or wrote comments on sticky notes and attached them on book pages and later transcribed them using word-processing program. Organizing notes was also an essential part of reading and note making.

4.2.2 Collecting and processing data

Historians' goal was to collect and process data for analysis. If primary sources existed only in physical form, they had to be digitized first. Some of the historians digitized the data themselves. They had visited in archives and taken pictures of primary sources with digital cameras. In some cases, the primary sources were already in digital form and accessed or downloaded using digital archives.

Digital primary sources sometimes needed further processing in order to be analyzed, e.g., by optical character recognition (OCR) or transcribing digital handwritten documents. In some cases, digital archive provided materials both in digital images and in OCR-format.

Two of the historians constructed databases manually. They first retrieved the digital primary sources either from their own collections or from digital archives. Then they started reading through the digital documents and collected information items into databases that they had created using a spreadsheet or a database management program. They also added new fields into databases when necessary.

Now I'm going to continue building this database [...] I insert here my original sources. And then I have created here a template, so I can later find the results easier and process them and make comparisons. (P5, shadowing)

Historian P1 was building a database from digital primary sources using programming (e.g. "R"). P1 wrote down the steps needed in performing the task, either to sticky notes or to some digital document. From time to time P1 used Google to solve the problems that arose during the programming and often ended up on Q&A programming websites and found answers quickly.

Historians' goal was to collect data in such a way that the information loss on the relevant data would be minimal. Historian P3 told that it was important to see whether the page numbers were in Roman or Arabic numerals. For another example, historian P5 wrote words to a database in their original language, when it was difficult to come up with a suitable translation. Historians also collected "address information" (e.g. web addresses or headlines of digital primary sources) to locate the origin of data fragments. This information was useful if they needed to go back to the original primary sources.

Some of the historians worked a lot with OCR-texts. Although the OCR-text was very useful, it caused a lot of extra work during data collecting and processing.

[...] it is worth checking out the actual OCR from the original source, because they often go page by page, and it might be that the article continues to the next page, and if you want the whole OCR from it.
(P4, interview)

Sometimes OCR-text contained misidentified characters, and sometimes words or paragraphs had switched places. For example, while building a database, historian P4 had to compare a digital image and its OCR-text with each other and correct the OCR-text. Historian P2 thought that it would be useful if for example editorials or advertisements could be separated from the rest of the data automatically.

Some archives offered primary sources as downloadable data packages. Historians considered this very useful. If data packages were not available, the digital documents had to be retrieved from the archives one by one. This made the use of primary materials difficult. Historians generally used e.g., txt- and csv-files. Sometimes files had to be named in a certain way in order to be used as input to the programs. Historian P2 processed the data for computer-aided analysis programs and at the same time tried to figure out the fastest ways to do the processing.

Hopefully [the digital primary sources] will be in such a format that they can be analyzed with [computer-aided] programs that I can do the actual analysis. This is like preprocessing. (P2, shadowing)

At least two historians were planning to deposit the data to an open access data archive for future use. One historian wrote a description of the dataset, for example, how data was collected and processed, and what possible problems it contained. Writing the description was considered important, because otherwise it would be hard for other researchers to use the data.

4.2.3 Analyzing

The main goal for analysis was understanding the phenomenon, although historians had more specific goals for analysis relating to their own studies. One way of understanding the phenomenon was asking questions. Historian P4 wrote down questions that arose while interacting with primary sources.

I'm sort of writing down the questions for myself and wondering if this could relate to this thing or not. It is kind of thinking aloud. [...] sometimes the questions that arise aren't relevant. But sometimes they are the ones that push the research forward. (P4, interview)

Some historians used or planned to use computer-aided programs in analyzing big data by calculating (e.g. word frequencies) or visualizing (e.g. geographical locations in a map). However, the work processes were not always ideal, e.g., an analysis program processed data longer than expected or did not provide the data in preferred format. In addition, poor OCR quality was a problem during the analyzing. One historian considered using alternative data because of the poor OCR quality. Especially when using big data, historians had to

evaluate the impact of the noisy OCR on the results and if the analysis is worth doing.

When asked whether they primarily used traditional or computer-aided research methods, historian P5 mentioned using only traditional methods during the research process, although P5 used some digital tools in analysis (e.g. spreadsheet) to make it easier to do comparisons and graphs. Historian P1 described using primarily computer-aided methods because of the vast amount of coding during the research process. Most historians combined traditional and computer-aided methods. Historian P4 described using computer-aided research methods during research process as follows:

On some parts, computer-aided methods are being used, but the whole process isn't algorithmic so that we could do a bigger analysis just on algorithms and could continue on making a historical analysis, but actually we use them in certain points of the work process. (P4, interview)

Historians worked to interpret the results. When the computer-aided analysis is done and the data is visible on the screen, the actual interpretation is based on the context of the data. Historian P2 described using analysis programs and making interpretations as follows:

When talking about digital history, the best presentations that I have seen have not been that miraculous in a methodological sense. They have been [...] for example calculating words. But then what has been great, is how the researcher has explained the contexts [...]. (P2, shadowing)

Seeing the original context was important also in the analysis stage. Therefore, historians went back to the primary sources for close reading. One historian explained that it was important to see if a news article was on the first or the last page of a newspaper, or what the other articles told about the era. Close reading was also a way to find quotations. Primary sources were retrieved for close reading from their own collections or from digital archives. In case of primary sources being big data, there was considerably less need for close reading but it was still a fundamental stage in the research process.

4.3 Synthesizing and Reporting

The main goal for writing was to produce a monograph, an article or a conference paper. Historians had sub goals, e.g., writing one chapter to a monograph or find a better focus for an article. Historians' writing was also connected to analyzing.

I feel like there is still a lot to do in terms of finding the final focus for the article. [...] But on the other hand, the focus cannot be reached until you have gone through these [primary sources] and then read more literature. Now I'm like in this vague state in this regard. (P4, shadowing)

But what I'm doing right now and what I should be doing is analyzing. In other words, to write, write and write. (P2, interview)

Text style, publication channel and target audience affected on what kind of texts the historians were writing. For example, one historian had a goal to write an international scientific article.

Historian P5 was working on with a current draft in order to shorten it and make it more readable. Historians had also time-related goals, for example, when the texts were supposed to be finished.

Historians writing styles varied. For example, historian P5 had a straightforward writing style, whereas historian P4 processed text little by little towards the final draft. Historian P1 started by drafting a general framework of the text which helped keeping the writing process on the right track.

[...] I will make a kind of a plan about the whole structure and then I'll start writing, and then of course it changes along the way. (P1, interview)

In the case of two historians (P2 and P4), the writing of the first version of a text seemed similar. Historian P2 described the writing process as follows:

[...] you write the first version, look all the notes that you have related to the topic from the notebooks and other notes. And maybe the preliminary results. And you put them all to the same Word-file. And then you check what you have. And then you start arranging what things are related to each other. And when you have written the first version, it's useful to take comments. (P2, interview)

Similarly, P4 combined different texts into an article draft and added citations from a reference management program. P4 told that, at first, the text file grows enormous but eventually the text narrows down as the focus of the text clarifies.

Actions related to writing were also locating the information needed for writing and using writing-support tools. Historians searched for information (e.g. notes, secondary sources, primary sources) from their own collections by browsing the file structures on their computers or using computers' own search fields. Historians made searches also from individual documents' text content. They searched information from reference management software, note-taking software and from library databases. For example, historian P5 used a note-taking software as a sort of a road map in order to locate the right document from personal collections.

[...] I have rarely put all my notes from the sources to [the note-taking software]. I have the longer versions of the notes in different files, but eventually [the note-taking software] is the help finding the right Word document. (P5, interview)

While processing big text files, it was important for one historian to be able to move quickly to a specific part of the text. Historian P4 made use of the headlining of the Microsoft Word text processing program.

[...] with the help of these [headlines] it is sort of easier to go from a part of the text to another when you are moving the text blocks into their right places. (P4, shadowing)

Reference management was an integral part of the writing process for historians. Some historians used reference management programs and some historians also maintained notes on the things they had not written yet on their texts. They updated these notes during the writing for example by putting a check mark next to the note they had already added.

4.4 Summary of Results

Altogether, we identified five task process types that were included in the following activities in the TBII model: searching and selecting information items, working with information items, and synthesizing and reporting. Table 5 shows the TBII activities on the left column, and the task process types with related goals and actions are on the right.

Table 5: Goals and actions in different task process types

SEARCHING AND SELECTING INFORMATION ITEMS	<p>SEARCHING SECONDARY SOURCES</p> <p>Goals: Acquire secondary sources (e.g. books and articles) Gaining a better focus for a writing task</p> <p>Actions: Using information retrieval systems Following citations Having conversations with colleagues Browsing library books Keeping an eye on interesting literature</p>
WORKING WITH INFORMATION ITEMS	<p>READING AND MAKING NOTES</p> <p>Goals: Learning (e.g. improving the basic understanding of the current topic) Making notes for future use</p> <p>Actions: Reading Following citations Making notes Organizing notes and citations</p> <p>COLLECTING AND PROCESSING DATA</p> <p>Goals: Collecting and processing data for analysis Depositing data to an open access archive</p> <p>Actions: Digitizing primary sources Building databases Preparing data for analysis Writing a description of a dataset</p> <p>ANALYSING</p> <p>Goals: Understanding phenomenon (e.g. describing it, comparing, detecting similarities and differences)</p> <p>Actions: Writing down questions Using analysis programs Interpreting data from databases or from results obtained using analysis programs Close reading of digital primary sources</p>
SYNTHESIZING AND REPORTING	<p>WRITING</p> <p>Goals: Finishing a writing work (main goal) Finding a focus for an article Writing for certain audience and forum Scheduling the writing process</p> <p>Actions: Writing and processing text Locating information items for writing Using tools to support writing</p>

5 DISCUSSION

In this research, we used TBII model [20] as a framework for task process types. We studied information interaction of historians with the focus on the use of digital primary sources. Further, we examined task goals and actions during the real-life task processes by interviewing and shadowing the participants.

The task process types are similar to the scholarly primitives in [32]. The action types during the task process type Searching secondary sources corroborate the findings in previous studies [11], although many of them concentrated on finding primary sources, not necessarily their use [1, 9, 31].

In case of the secondary sources, we found that some historians felt that following citations was a better way of finding them than using information retrieval systems. Citations provide context to the research and help finding information efficiently. This is similar to Bates' *footnote chasing* and *citation searching* [2]. This action was also important during reading and making notes. Therefore, it should be as smooth as possible. Links across information items is already an existing feature of exploratory search functionalities in some digital libraries [17]. However, the links should cover also different types of information allowing citation following from primary to secondary sources. This means integration of heterogeneous information resources [22]. In the future, time spent reading in digital environments will increase and therefore it requires more attention from system designers [15].

Whereas following citations was a principal action in searching secondary sources, historians also considered IR systems useful. In our study, IR systems were used to find a better focus for a writing task. This is an example of an exploratory search, where the goal is about learning rather than fact-finding [23]. However, the existing IR interfaces show results in lists and do not adequately support exploration. Systems helping to understand connections between information items, e.g., by topical exploration as suggested by [29], may better support this action. However, this support should also be expanded across varying types of sources.

It became evident that the information needs in search tasks are more versatile and dynamic than the systems expect them to be. For example, one historian searched literature to support the writing process. While the main goal of finding useful secondary sources remained the same, there were changes in the sub goals during the task. In some points, literature was searched with a sharp focus to find a helpful reference, and occasionally, the aim was to find literature more generally. We did not study this in detail because our focus was broader and its study requires more focused research setting. However, the changes in sub goals described above are similar to shifts of interactive intentions presented by Xie [34]. Historians lack support from IR interfaces in expressing their information needs. Huurdeman [19] suggested that at the beginning of a writing process people need low-level support such as input and control features, e.g., word clouds. Yet we found that even the simplest input or control features were missing. For example, one historian noted that

some IR interfaces lacked possibilities to narrow down the search to the sub-field of history or to a specific historical era.

Further, we found that motivations in accessing the information resources were depended on the task process type. In case of the primary sources, historians preferred to download data packages when collecting and processing data. By downloading bulks of data, they saved time instead of downloading items one by one. On the other hand, in the analysis phase they wanted to examine individual documents in detail to confirm the context of the data, and retrieved the documents for close reading. These kinds of different uses of information items should be designed into digital archives and tools to help the historians work tasks.

Historians wanted to save time during collecting and processing data because of time consuming analyzing and writing phase. As Given and Willson [18] noted, data preparation phase is often overlooked when talking about research practices of digital humanities researchers. Because of its time-consuming nature, instead of just providing digital primary sources to be clicked and opened one by one, the data should be accessible also as data packages and in different formats.

Similar to Given and Willson's [18] study, we discovered that historians used different kinds of tools to support their writing, e.g. note-taking or reference management software. In addition, we noticed that writing was a part of analyzing, a way of making connections and developing new ideas. Accordingly, Brockman et al. [6] found humanities scholars' writing as a uniting activity for many other activities e.g., reading and studying primary sources.

Considering the limitations of this study, the presence of the researcher may have affected how the historians worked during the shadowing, e.g., they may have selected simpler and more focused task and not all possible task process types are included. Furthermore, the shadowing was limited to situations where the historians were working alone at their computers. However, it became evident that they worked in co-operation with other researchers. Further historians conducted self-chosen real tasks during the shadowing. Although we studied few historians' performance and the results may not be generalizable to larger population, we believe that observing real world situations give rich view on information interactions that cannot be captured by any other means.

6 CONCLUSION

The purpose of this research was to find out what is historians' information interaction like in digital environments. As a conclusion, we argue that historians need better support for expressing their information needs in information retrieval systems. Furthermore, the IR systems need features that support exploratory searching. The task processes have an effect on how the digital tools are being used. This should be noted when designing digital tools for historians. Further, data processing is a time-consuming task, which could be supported better with automated tools and digital archive design.

ACKNOWLEDGMENTS

This research was supported by the Academy of Finland grant number 310278.

REFERENCES

- [1] I. G. Anderson. 2004. Are you being served? Historians and the search for primary sources. *Archivaria*, 58(Fall 2004), 81-129.
- [2] M. J. Bates. 1990. Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5), 575-591.
- [3] M. J. Bates, D. N. Wilde and S. Siegfried. 1993. An analysis of search terminology used by humanities scholars: The Getty Online Searching Project Report Number 1. *The Library Quarterly*, 63(1), 1-39.
- [4] A. Blandford and S. Attfield. 2010. *Interacting with information*. Morgan & Claypool, San Rafael.
- [5] C. L. Borgman. 2007. *Scholarship in the digital age: information, infrastructure, and the Internet*. MIT Press, Cambridge, Mass.
- [6] W. S. Brockman, L. Neumann, C. L. Palmer and T. J. Tidline. 2001. Scholarly work in the humanities and the evolving information environment. Digital Library Federation. Council on Library and Information Resources. Washington, D.C. Retrieved from <https://www.clir.org/pubs/reports/pub104/>.
- [7] A. Brundage. 2017. *Going to the sources: A guide to historical research and writing*. (6th ed.). John Wiley & Sons, Incorporated, Hoboken, NJ.
- [8] K. Byström and P. Hansen. 2005. Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology*, 56(10), 1050-1061. DOI: 10.1002/asi.20197.
- [9] A. Chassanoff. 2013. Historians and the use of primary source materials in the digital age. *The American Archivist*, 76(2), 458-480. DOI: [//doi.org/10.17723/aarc.76.2.lh76217m2m376n28](https://doi.org/10.17723/aarc.76.2.lh76217m2m376n28).
- [10] J. Cullen. 2017. *Essaying the past: How to read, write, and think about history*. (3rd ed.). John Wiley & Sons, Incorporated, Chichester, UK; Hoboken, NJ.
- [11] M. S. Dalton and L. Charnigo. 2004. Historians and their information sources. *College & Research Libraries*, 65(5), 400-425. DOI: 10.5860/crl.65.5.400.
- [12] W. M. Duff and C. A. Johnson. 2002. Accidentally found on purpose: Information-seeking behavior of historians in archives. *The Library Quarterly*, 72(4), 472-496. DOI: 10.1086/lq.72.4.40039793.
- [13] D. Ellis. 1993. Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4), 469-486. DOI: 10.1086/602622.
- [14] A. T. Erickson. 2013. Historical research and the problem of categories: reflections on 10,000 digital note cards. In K. Nawrotzki and J. Dougherty (Eds). *Writing history in the digital age*. University of Michigan Press, Ann Arbor, MI. DOI: [//dx.doi.org/10.3998/dh.12230987.0001.001](https://dx.doi.org/10.3998/dh.12230987.0001.001).
- [15] L. Freund, R. Kopak and H. O'Brien. 2016. The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science*, 42(1), 79-93. DOI: 10.1177/0165551515614472
- [16] L. Freund and E. G. Toms. 2016. Interacting with archival finding aids. *Journal of the Association for Information Science and Technology*, 67(4), 994-1008. DOI: 10.1002/asi.23436.
- [17] N. Fuhr, C. Klas, A. Schaefer and P. Mutschke. 2002. Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In M. Agosti and C. Thanos (Eds). *Research and Advanced Technology for Digital Libraries*. (6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002). Springer, Berlin, Heidelberg, 597-612.
- [18] L. M. Given and R. Willson. 2018. Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6), 807-819. DOI: 10.1002/asi.24008.
- [19] H. C. Huurdeman. 2018. Supporting the complex dynamics of the information seeking process. Ph.D. Dissertation. University of Amsterdam, The Netherlands. Retrieved from <http://hdl.handle.net/11245.1/1e3bf31a-0833-4ead-a00c-4cb1399d0216>.
- [20] K. Järvelin, P. Vakkari, P. Arvola, F. Baskaya, A. Järvelin, J. Kekäläinen, H. Keskustalo, S. Kumpulainen, M. Saastamoinen, R. Savolainen and E. Sormunen. 2015. Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems*, 33, 1, Article 3 (March 2015), 30 pages. DOI: [//dx.doi.org/10.1145/2699660](https://doi.org/10.1145/2699660).
- [21] C. C. Kuhlthau. 1993. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339-355. DOI: [//doi.org/10.1108/eb026918](https://doi.org/10.1108/eb026918).
- [22] S. Kumpulainen and K. Järvelin. 2010. Information interaction in molecular medicine: integrated use of multiple channels. In N. Belkin & al. (Eds.), *Proceedings of the Information Interaction in Context conference (IIIX 2010)*, New Brunswick, NJ, August 2010. ACM, New York, NY, USA (pp. 95-104.) DOI: 10.1145/1840784.1840800.
- [23] G. Marchionini. 2006. Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4), 41-46. DOI: 10.1145/1121949.1121979.
- [24] K. Martin and A. Quan-Haase. 2013. Are e-books replacing print books? Tradition, serendipity, and opportunity in the adoption and use of e-books for historical research and teaching. *Journal of the American Society for Information Science and Technology*, 64(5), 1016-1028. DOI: 10.1002/asi.22801.
- [25] K. Martin and A. Quan-Haase. 2016. The role of agency in historians' experiences of serendipity in physical and digital information environments. *Journal of Documentation*, 72(6), 1008-1026. DOI: 10.1108/JD-11-2015-0144.
- [26] K. Martin and A. Quan-Haase. 2017. "A process of controlled serendipity": An exploratory study of historians' and digital historians' experiences of serendipity in digital environments. In S. Erdelez & N. K. Agarwal (Eds.), *Proceedings of the Association for Information Science and Technology* (pp. 289-297.) Hoboken, NJ: Wiley. DOI: 10.1002/pra2.2017.14505401032.
- [27] S. McDonald. 2005. Studying actions in context: a qualitative shadowing method for organizational research. *Qualitative research*, 5(4), 455-473. DOI: 10.1177/1468794105056923.
- [28] S. McDonald and B. Simpson. 2014. Shadowing research in organizations: the methodological debates. *Qualitative Research in Organizations and Management: An International Journal*, 9(1), 3-20. DOI: 10.1108/QROM-02-2014-1204.
- [29] T. Ruotsalo, G. Jacucci, P. Myllymäki and S. Kaski. 2015. Interactive intent modeling: Information discovery beyond search. *Communications of the ACM*, 58(1), 86-92. DOI: 10.1145/2656334.
- [30] J. Rutner and R. C. Schonfeld. 2012. Supporting the changing research practices of historians. Final Report from ITHAKA S+R. DOI: [//doi.org/10.18665/sr.22532](https://doi.org/10.18665/sr.22532).
- [31] H. R. Tibbo. 2003. Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *The American Archivist*, 66(Spring/Summer 2003), 9-50.
- [32] J. Unsworth. 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this. *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. (King's College, London, May 13, 2000).
- [33] R. Vuorikari, Y. Punie, S. Carretero Gomez and G. Van den Brande. 2016. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model. Luxembourg Publication Office of the European Union. EUR 27948 EN. DOI: 10.2791/11517. Retrieved from <https://publications.europa.eu/en/publication-detail/-/publication/bc52328b-294e-11e6-b616-01aa75ed71a1/language-en#>.
- [34] H. Xie. 2000. Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science*, 51(9), 841-857.