DHQ: Digital Humanities Quarterly

2016 Volume 10 Number 4

Information access in the art history domain: Evaluating a federated search engine for Rembrandt research

Suzan Verberne <s_dot_verberne_at_let_dot_ru_dot_nl>, Radboud University Lou Boves <l_dot_boves_at_let_dot_ru_dot_nl>, Radboud University Antal van den Bosch <a_dot_vandenbosch_at_let_dot_ru_dot_nl>, Radboud University

Abstract

The art history domain is an interesting case for search engines tailored to the digital humanities, because the domain involves different types of sources (primary and secondary; text and images). One example of an art history search engine is RemBench, which provides access to information in four different databases related to the life and works of Rembrandt van Rijn. In the current paper, RemBench serves as a case to (1) discover the requirements for a search engine that is geared towards the art history domain and (2) make recommendations for the design of user observation studies for evaluating the usability of a search engine in the art history domain, and in digital humanities at large.

A user observation study with nine participants confirms that the combination of different source types is crucial in the art history domain. With respect to the user interface, both free-text search and facet filtering are actively used by the observed participants but we observe strong individual preferences. Our key recommendation for specialized search engines is the use of faceted search (free text search combined with filtering) in combination with federated search (combining multiple resources behind one interface). In addition, the user study shows that the usability of domain-specific search engines can successfully be evaluated using a thinking-aloud protocol with a small number of participants.

1. Introduction

For the interpretation of works of art, art historians study these works along with different types of textual sources: primary documents describing the works of art, their maker, the patron, the provenance, etc., but also contemporary documents related to the social and economic context in which the artist worked, and secondary literature. The information about works of art, both as primary and secondary documents, can be found independently in digital (and in part also non-digital) resources produced by museums, archives and libraries. Increasing amounts of sources residing in libraries and archives are digitized and made accessible online [Rodríguez Ortega 2013].

1

2

In a previous project ("RemDoc") we developed a digital collection of primary documents that relate to the life and works of Rembrandt van Rijn (1606-1669). These documents had partly been published before in the form of books [Strauss and van der Meulen 1979] [Roscam Abbing 2006]. In the Rembrandt Documents ("RemDoc") project we digitized them, added metadata, and made them available through an online search engine.^[1] Digital collections such as RemDoc form a pivotal tool for the research of art historians: they serve as the entry point for and guidance in research projects. However, for many of the research questions faced by Rembrandt researchers, often other sources than primary documents are needed. Here are a few examples of questions that might be asked by Rembrandt researchers:

- Was Rembrandt's "Reading Woman" in the Rijksmuseum painted on canvas or panel?
- Did Rembrandt paint dogs?
- How old was Titus when he died?
- Where is Rembrandt's "Storm on the Sea of Galilee"?

· Find etchings after Rembrandt's self portraits

To answer these questions, researchers need to combine diverse and geographically distributed sources of information: text, images, and bibliographical data. Probably the most consulted resources for art historical information on the topic of Rembrandt are the databases of the Netherlands Institute for Art History (RKD). The database RKDimages contains descriptions and images of mainly Dutch paintings, drawings, prints and original photos before 1940; the database RKDartists contains biographical information about Dutch and foreign artists of the middle ages to the present. For Rembrandt research, the RKD databases are an important reference. In addition, secondary literature stored in libraries is relevant for art history research. The dispersed nature of relevant resources in the Rembrandt research domain motivated us to design and develop an online work environment that integrates these resources.

3

4

5

6

This resulted in the development of RemBench.^[2] RemBench provides access to information related to the life and works of Rembrandt van Rijn stored in four different databases: RemDoc, RKDimages, RKDartists, and a university library catalogue. These databases are made accessible through a single search interface. The target user groups of RemBench are art historians, historians, and other humanities scholars and students who are interested in Rembrandt's period, the 17th century. The goal of the RemBench project was to show the value of online digital data for art history research. We defined two sub-goals: (1) to connect the metadata of four different databases and (2) to build a demonstrator search interface to these databases. The focus of the current paper is on the development and evaluation of a search interface that gives access to the four databases.

Access to the four different databases is realized through *federated search*. Federated search (sometimes called *distributed information retrieval*) is a technique for searching multiple collections simultaneously with a single query. The results returned by the selected collections are integrated into a single result page [Jascó 2004] [Shokouhi and Si 2011]. Federated search is often used in online library services that give access to publications stored in multiple different repositories, but most people also know it from the use of Google, where one query often generates results from Google images, Google maps and the textual Google index. In the terminology of federated search, the different resources are called *verticals*.

Because of the importance of metadata for searching in the art history domain [Bates 1996a] [Yee, Swearingen, Li, and Hearst 2003], we decided that RemBench should not only allow for free text search, but also for filtering metadata values. For example, filtering on location can help find works of art that are stored in Moscow; filtering on document type can help find birth certificates, and filtering on date can help find documents from before 1700. The search engine technology that combines free text search with filtering for metadata values is called *faceted search* [Tunkelang 2009]. Faceted search is commonly used in web shops, booking sites and review sites, where users can select options for the items they are searching for. Figure 1 shows the faceted search interface of RemBench. RemBench will be described in more detail in Section 3.



For the development and evaluation of domain-specific search engines, it is important to know the information-seeking behaviour of the target group [Gossen, Hempel, and Nürnberger 2013]: what type of queries do they phrase given a research question, how do they formulate follow-up queries, and how do they decide which results are relevant? However, these behaviours are to a large extent determined by the functionality and usability of the search system at hand. At the same time, the satisfaction of a user group with a search system depends on the match between the functionality and design of the system and the goals of the users. This interaction makes it difficult to disentangle information-seeking behaviour and the requirements that a search engine should meet. It is therefore crucial to observe the target group in interaction with a search engine in order to understand their information-seeking behaviour and evaluate the usability of the search engine. Therefore, we set up a user observation study in order to address the following research questions, using RemBench as a case study:

7

8

9

1. What are the requirements for a search engine that is geared towards the art history domain?

These requirements follow from the analysis of three user aspects: (a) the questions asked by Rembrandt researchers; (b) the user-system interactions that we observe, and (c) the measured user satisfaction. In Section 5, we will address these three aspects.

2. What are recommendations for the design of a user observation study for evaluating the usability of a search engine in the art history domain?

These recommendations follow from our experiences with the user observation study, combined with the recommendations from earlier work on the usability of search systems.

In Section 2 we discuss the relevant literature. In Section 3 we describe the RemBench project. In Section 4 we present the methodology for our user observation study, followed by results and analysis in Section 5. In Section 6 we answer the two research questions, identify limitations of the user study and make recommendations. In Section 7 we present our conclusions.

2. Literature review

For answering our research questions, two topics from the literature are important: the information-seeking behaviour of humanities scholars (Section 2.1) and the methodology for evaluating the usability of domain-specific search systems and analysing user-system interactions (Section 2.2).

2.1. The information seeking behaviour of humanities scholars

When designing search systems for specific target groups it is important to take into account the information-seeking behaviour of the target user group [Gossen, Hempel, and Nürnberger 2013]. Many studies address the information-seeking behaviour of specific target groups, e.g. of medical doctors [Davies 2007], social scientists [Meho and Tibbo 2003], academic scientists in general [Hemminger, Lu, Vaughan, and Adams 2007], graduate students [George, Bright, Hurlbert, Linke, St Clair, and Stein 2006], lawyers [Makri, Blandford, and Cox 2008] and doctoral students in the field of biology [Vezzosi 2009]. Some studies address the information-seeking behaviour of humanities scholars, or even more specifically, of art historians. For the current research it is relevant to know whether humanities scholars – and art historians in particular – have different information-seeking behaviour than other groups and therefore have different needs for domain-specific search engines.

The work by Bates et al. [Bates, Wilde, and Siegfried 1993] [Siegfried, Bates, and Wilde 1993] [Bates, Wilde, and Siegfried 1996] [Bates, conducted over a two-year period. Complete logs of the searches and output were captured, and a small number of scholars were interviewed in depth. Based on these interviews, it is argued [Bates, Wilde, and Siegfried 1994] that humanities scholars have different needs than researchers from other fields, because (a) publications in the humanities are not so much about fact discovery, but more about the author's opinions, and (b) humanities scholars use very different terms than researchers in other fields to describe the information they need. However, the study was conducted in a time when web-based search engines were still in their infancy, which might mean that the reported behaviour [Siegfried, Bates, and Wilde 1993] may not be representative of current information seeking behaviour. Nevertheless, some of the recommendations made in the final report [Bates 1996a] still hold, such as the recommendation to combine multiple databases in one search engine, and to enable faceted search:

"Recognize that humanities searches are often composed of several facet elements. Because the indexing for many databases is not optimally designed for queries of this sort, good online database searching in the humanities may actually be harder than for the sciences, even for the skilled online intermediary, and will almost certainly be difficult for the typical humanities end user." [Bates 1996a, 9]

It could be argued that the difficulty mentioned in the last sentence does not hold anymore nowadays, since all end users are expected to have experience with web search engines and web interfaces to databases [Rowlands, Nicholas, Williams, Huntington, Fieldhouse, Gunter, ... and Tenopir 2008]. For example, it was found that the expectations and behaviour of the users of digital libraries have changed since the mid-90s [Sadeh 2007]. In 2007, Hemminger et al. found that although bibliographic and citation databases still were the most commonly used tools by academic scientists for finding information, these were closely followed by search engines and digital libraries. Hemminger et al. note that "As free, Web-based literature databases such as Google Scholar continue to grow, the distinction between bibliographic/citation databases and Web search engines is blurring." [Hemminger, Lu, Vaughan, and Adams 2007, 2210]. Around the same time, George et al. [George, Bright, Hurlbert, Linke, St Clair, and Stein 2006] found that 97% of graduate students use the Internet (non-library websites) as source for information and 73% report the use of Google for their research-related information seeking. This is also the case for art historians, as has been shown in

12

15

more recent research by Beaudoin and Brady [Beaudoin and Brady 2011], who studied the search behaviour of archaeologists, architects, art historians and artists in digital image libraries. One of their findings was that the art historians who participated in their study all reported Google Images as their favourite resource for images.

On the other hand, the study by George et al. [George, Bright, Hurlbert, Linke, St Clair, and Stein 2006] shows that the use of Google is the least frequent for humanities students, compared to students from other fields, while their use of library databases is the highest compared to other fields. This may be explained by the importance of primary sources for the humanities, which often reside in libraries and archives. Tenopir et al. [Tenopir, King, Spencer, and Wu 2009] confirm this. They found that humanities scholars read fewer articles than academics from other disciplines, and that they rely more on browsing (as opposed to searching, using citations and asking colleagues) for finding relevant articles. However, "this is not to say that humanities faculty members do not read, but they most likely read books, primary materials, and manuscripts" [Tenopir, King, Spencer, and Wu 2009, 147].

17

20

21

The effectiveness of faceted search for research in the art history domain has been shown by Yee et al. [Yee, Swearingen, Li, and Hearst 2003], who present the results of a usability study in which art history students explored a collection of 35,000 fine arts images, using a faceted interface for metadata filtering. They find that for these data, 90% of the participants preferred the metadata approach over a free-text search functionality, despite the fact that the latter was more familiar to them. This was likely due to the nature of the data: images cannot easily be disclosed through free text search. Metadata can disclose images, but the user then depends on the terms chosen by the developers of the database for labelling the images. If there is a mismatch (or the users suspect a mismatch) between the terms in the database and the terms they would use themselves, metadata filtering is a better option than free-text search, because selecting the most likely (best matching) term(s) from a list is more effective than guessing query terms.

In summary, the literature shows that humanities scholars are experienced in the use of web search engines, but rely on digital libraries more than researchers in other fields. Because of the importance of metadata in the art history domain, faceted search is an important feature for the interface to databases and libraries.

2.2. Usability testing and user interaction analysis for information seeking

It has been shown that there is a strong relationship between usability and popularity of web search engines [Dudek, Mastora, and Landoni 2007]; thus it is important that usability is evaluated properly. The most-used method for usability testing is to have users from the target group work with the software or website, and ask them to provide feedback on the usability. With respect to the number of users needed, Nielsen and Landauer [Nielsen and Landauer 1993] show that for most usability tests, the proportion of additional usability problems found when adding test users quickly decreases beyond five users. A common method for registering user feedback is *thinking aloud*. Van Waes [Van Waes 2000] found that for usability testing of web applications, the combination of thinking aloud with recording of the navigation process is a valuable observation method, "to collect data about various usability aspects, navigation strategies, and cognitive aspects of the searching process" [Van Waes 2000, 288]. The common way of evaluating the usability of search engines is to give users a series of information problems and ask them to find the answers using the search engine at hand. Afterwards, the users are presented with a questionnaire in order to assess their satisfaction with the system [Spink 2002]. This latter method of usability testing is adopted in the current paper; in addition, we combine screen capturing with thinking-aloud as suggested by Van Waes [Van Waes 2000] to collect the user interactions.

In the interaction between a user and a search system, two aspects play a central role: (1) query formulation and reformulation and (2) the examination and judgement of search results. Both topics have been addressed before:

 Query formulation and reformulation strategies in multi-query sessions are described by Lau and Horvitz [Lau and Horvitz 1999] and Rieh [Rieh 2006]. They distinguish query specification (making the query narrower by adding one or more terms), query generalization (making the query broader by removing one or more terms) and query reformulation (replacing one or more terms by other terms without making the query broader or narrower). The effect of query specification is generally a higher precision and lower recall of relevant results in the result set, while the effect of query generalization is the opposite. In the case of faceted search interfaces, query formulation does not only comprise the typing of textual queries, but also the selection of facet values. Kules et al. [Kules, Capra, Banta, and Sierra 2009] found in an eye-tracking study that users of a faceted search interface working on an exploratory search task spent about 50 seconds per task looking at the results, about 25 seconds looking at the facets, and only about 6 seconds looking at the query itself. These results suggest that facets are important for users with exploratory search tasks.

• For investigating how users process and evaluate the items returned in response to a query, important work has been done with eye-tracking experiments [Granka, Joachims, and Gay 2004]. One of the findings of this work is that users spend much more time looking at the results in the top five of the result list (especially the first and second result) than at the results in position six to ten. In the case of federated search, search results come from multiple information sources (so-called "verticals") and are presented to the user in one result page. It is interesting to study how users interact with the combined result page [Ponnuswami, Pattabiraman, Wu, Gilad-Bachrach, and Kanungo 2011] and what the differences are in user behaviour for results from the different verticals [Spink and Jansen 2006]. Ponnuswami et al. [Ponnuswami, Pattabiraman, Wu, Gilad-Bachrach, and Kanungo 2011] address the challenge of "page composition", where the goal is to create a combined result page with the most relevant information from the different verticals.

Query reformulation patterns will be addressed again in Section 5.2.d, and result-page composition in Section 3.

3. About RemBench

RemBench connects four existing databases. The first two are RKDartists and RKDimages, two art historical databases maintained by the Netherlands Institute for Art History (RKD)^[3]. RKDartists is a database of biographical information about Dutch and foreign artists from the middle ages to the present day. RKDimages is a database with descriptions and images of mainly Dutch paintings, drawings, prints and original photos from before the Second World War. For RemBench, a selection of records from the RKD databases was made that only includes artists and images that are related to Rembrandt van Rijn. The starting point for this selection was the record of Rembrandt himself in RKDartists. In the selection of artists, all artists were included that are mentioned in the Rembrandt record (either as "pupil of", "teacher of", "followed by", "influenced by" or "had influence on"). In the selection of images, all works of art from RKDimages were included that have been created by one of the artists in the artist selection. The resulting selection consists of 1,857 works of art and 59 artists.

The third database connected to RemBench is RemDoc,^[4] the digital collection of primary documents that relate to the life and works of Rembrandt van Rijn. In the RemDoc project, all known documents have been collected and disclosed that relate to Rembrandt, as a person and as an artist, as well as to his ancestors, family, offspring and business partners, from the 15th to 18th centuries. The database contains 2,003 documents. Since all these are related to Rembrandt, all were included in RemBench.

The fourth database in RemBench is a university library catalogue, digitized and disclosed through the RUQuest^[5] system, a library search system that provides access to the full collection of the Radboud University Library, and full-text articles of all journals that Radboud University subscribes to. From the data available through RUQuest, a selection was made that is defined by all records that are returned for the query "Rembrandt" (68,958 records in total).

The RemBench project consisted of two phases: First, the metadata of the four databases were connected by mapping them onto one common metadata scheme. Second, a search engine was developed to disclose the data in these databases. The technology for faceted search and free text search that allows the user to search all databases at once was developed by the Huygens Institute for the History of the Netherlands using Apache Solr.^[6] The four databases were treated as four separate verticals in the interface, not merging the results in one result list. The main reason for that decision came from the art historians in the research team: They argued that four different groups of results would be more clear to the user. By doing so, the challenge of ranking the results from the different verticals relative to each other [Ponnuswami, Pattabiraman, Wu, Gilad-Bachrach, and Kanungo 2011] was avoided. In the user interface, the

22

verticals are labelled "Works of art" (group of results from RKDimages), "Artists" (group of results from RKDartists), "Primary Sources" (group of results from RemDoc), and "Library Sources" (group of results from RUQuest). In every vertical, maximally five results are shown on the front page. To see more results (in an overlay screen), the user has to click "More".

The RemBench user interface was designed by a professional designer, in interaction with art historians, who added their specific wishes. The beta version of RemBench was used in the user observation study described in the next section.

4. User study

As introduced in Section 1, we set up a user study for the observation of information-seeking behaviour by the target group of RemBench. Based on the observed information-seeking behaviour, we address the first research question: "What are the requirements for a search engine that is geared towards the art history domain?" For the user study, we adopt the methods for usability testing described in Section 2.2: we asked users from the target group to work with RemBench and to describe their actions and motivations using a *thinking aloud* protocol. We provided the subjects with a number of information problems and asked them to find the answers using RemBench. This set-up allows us to observe the information-seeking behaviour of RemBench users.

4.1. Design, materials and procedure

Students of history and art history (bachelor and master level) from Radboud University were recruited to participate in the user study; nine participants (2 male, 7 female, median age 20,5) in total. The participants were paid a volunteer fee of €10.

The task for the students was to find the best possible answers to a series of questions related to Rembrandt, using RemBench. Each participant was given 10 questions (see Section 4.2), one at a time. Some of the questions expected a single answer (yes/no, name, title, place), others a list of items. The participants were asked to use all functionalities of RemBench they needed to find the answers and stop their search when they felt that they had tried everything to find the answers. It was emphasized in the instructions that not their capabilities were not being tested, but the usability of RemBench.

The participants were working on a Windows 7 PC with Firefox. A researcher started the RemBench homepage for them and gave them the list of questions to work with.

User-system interaction was observed using a thinking-aloud set-up [Gerjets, Kammerer, and Werner 2011]: The participants were asked to voice their thinking process aloud, what actions they took in the search process and why they took them. Before the participants started working with RemBench, they were asked to orally describe the route they take from home to the university and count the number of traffic lights, as a way of familiarizing them with the thinking-aloud procedure [Van Waes 2000]. A researcher was sitting next to the participant and took extensive notes of what the participant did and said. The researcher tried to minimize the number of interruptions of the participant, but in some cases she requested the participant to go to the next question if the participant did not make any progress in finding the answer(s) to a question. Desktop activity was recorded using screen capture software,^[7] and camera recordings of the user and the computer were made from behind the user. After 45 minutes, the researcher asked the participant to finish the current question and skip the remaining questions.

After each question, the participants were asked to write down the answer they found on paper, and give two evaluative judgments:

- How satisfied are you with the answer found? (5-point rating scale)
- How satisfied are you with the use of RemBench for answering the question? (5-point rating scale)

After finishing all questions, the participants were given a post-task questionnaire with two evaluative questions:

32

27

- Please list the positive aspects of RemBench
- Please list the negative aspects of RemBench

4.2. Questions about Rembrandt

Two art historians from our research team (who are working in the Rembrandt domain themselves) formulated a number of questions about Rembrandt that are likely to be addressed by Rembrandt researchers. They provided 61 questions. The questions were formulated independently of the possibilities of RemBench as search tool; this implies that it is possible that for some of the questions the answer cannot be found using RemBench. With these 61 questions, and nine participants who were each given 10 questions, some of the questions could be offered twice, which made it possible to measure agreement between the participants.

4.3. Annotation of the data

The thinking-aloud data, the screen recordings and the notes taken by the researcher during the user observations were used afterwards to create an annotated data set of search sessions. A search session is defined as the series of all actions related to one question [Sriram, Shen, and Zhai 2004]. The sessions were manually annotated with the following information: the queries issued, the facets used, and the verticals used. In the case of facets, "using" means "selecting at least one value of the facet"; in the case of verticals, "using" means clicking at least one result from a vertical, or interpreting the results shown for a vertical as evidence for the answer.^[8] In the case that a user opened a result and clicked on the permalink^[9] to the original database (RKD, RemDoc or RUQuest), it was counted as an additional use of a vertical. Examples of the annotation are in Table 1.

Question	Facets used	Queries issued	Verticals used
How old was Titus when he died?		-titus	-Artists
Which Works by Rembrandt and his students have been in the possession of Abraham Bredius?	-Person type - Content type	-Abraham Bredius -Abraham Bredius possession -Bredius	-Artists -Works of art
Was Rembrandt's <i>Reading Woman</i> in the Rijksmuseum painted on canvas or panel?	- Content type - Location	-Lezende vrouw	-Works of Art -RKD

Table 1. Examples of user interactions for three questions

5. Results and analysis

In the next subsections, we analyse the following user aspects using the collected observation data:

- Question types: Which types of questions are asked by Rembrandt researchers? (Section 5.1);
- User-system interaction (Section 5.2):
 - · How many queries, facets and verticals do the users need to answer one question?
 - How does the session length for RemBench compare to library search and general web search?
 - What are the individual differences between users?
 - · How do users formulate and reformulate queries (query modification patterns)?
- User satisfaction (Section 5.3):
 - How does usability satisfaction relate to answer satisfaction?

37

35

- · How does user satisfaction differ between question types?
- What positive and negative comments did the users make?

5.1. Which types of questions are asked by Rembrandt researchers?

We classified the 61 questions phrased by the Rembrandt researchers in three types, using the question taxonomies of Hermjakob et al. [Hermjakob, Hovy, and Lin 2000] and Voorhees [Voorhees 2003]: closed questions, factoid questions, and list questions. The distribution of question types is shown in Table 2. Factoid questions are the most frequent, but there is also a relatively large share of closed questions and list questions. Note that list questions are the most difficult to answer because they require exhaustive searching: In order to answer a question starting with "which works are", the user has to find *all* words for which the description in the question is true. Factoid questions starting with *how many* have the same characteristic and can be equally difficult to answer.

Туре	Count	Examples (translated from Dutch)
Closed (yes/no; two options)	17	-Was Rembrandt's <i>Reading Woman</i> in the Rijksmuseum painted on canvas or panel? -Did Rembrandt know Jews? -Did Rembrandt paint dogs?
Factoid	28	-How old was Titus when he died? -How many works by Rembrandt are in private collections? -Where is Rembrandt's <i>Storm on the Sea of Galilee</i> ?
List	16	 -Which paintings by Rembrandt have been in the collections of the House of Orange-Nassau? -Which works by Rembrandt are in St. Petersburg? -Find etchings after Rembrandt's self portraits

Table 2. The questions classified by question type, with examples

5.2. User-system interaction

Fifty-four of the 61 questions were addressed by at least one participant. The other seven were skipped because not all participants succeeded in answering all 10 questions assigned to them in the 45-minute time slot. Fifteen questions were answered by two participants.

5.2.a. How many queries, facets, and verticals do the users need to answer one question?

From the annotated data (see Section 4.3), we extracted session statistics.^[10] The overall counts are given in Table 3. The results show that on average, a user issues two queries for a question, but the standard deviation is large. Figure 2 shows the number of interactions (facets, queries and source types) for the three question types in our data. A *t*-test for independent samples shows that the difference between the number of interactions for factoid questions and list questions is significant on the 0.01-level (p= 0.008): list questions require more user actions than factoid questions.

	Mean (stdev)	Min-max
Number of facets used	0.86 (1.00)	0-3
Number of queries used	2.11 (1.48)	0-7
Number of verticals used	1.46 (1.03)	0-5

 Table 3. The average numbers of facets, queries and verticals used per session

38

39



Figure 2. The number of facets, queries and verticals used, per question type

5.2.b. How does the session length for RemBench compare to library search and general web search?

We compare the session length for RemBench for session length in three other search logs:

- library search: data from The European Library (TEL) search logs, analysed by Verberne et al. 2010 [Verberne, Hinne, van der Heijden, Hoenkamp, Kraaij, van der Weide 2010]^[11]
- general web search in the early years of web search systems [Jansen, Spink, and Saracevic 2000]
- a more recent web search data set from 2006: the Microsoft Search (MSN) dataset that was distributed for the WSCD 2009 workshop,^[12] earlier analysed by Hinne et al. 2011 [Hinne, van der Heijden, Verberne, and Kraaij 2011]

In Figure 3 the distribution of numbers of queries in RemBench sessions is compared to the distribution for library search and web search. The average number of queries per session reported by Jansen et al. for the Excite search engine was 1.6; for the MSN data, the average number of queries per session is 1.5, which both are a bit lower than the 2.1 of RemBench. For the TEL data, the average number of queries per session is similar to RemBench: 2.0.

Figure 3 shows that especially the proportion of sessions with only one query differs between the different search contexts. The high proportion of sessions with one query can probably be explained by the high proportion of navigational queries on the web (33% of web queries, according to White and Drucker [White and Drucker 2007]), where the user already knows which web page (s)he wants to find. The proportion of one-query sessions increased between 2000 and 2006, probably because web search engines became more efficient and users more often found what they needed with one query. The proportion of single-query sessions in the TEL logs is a bit smaller than for web search, but larger than for RemBench.

41



set, Verberne et al. [Verberne, Hinne, van der Heijden, Hoenkamp, Kraaij, van der Weide 2010]) and for web search (the latter according to Jansen et al., [Jansen, Spink, and Saracevic 2000] and the MSN data set, 2006)

Another difference between RemBench, TEL search, and web search is that in the web search logs and the TEL logs there are no sessions without queries. In the RemBench interaction data, there are four sessions without text queries; in these cases, metadata filtering was sufficient to answer the question. Examples are the question "Are there works of Rembrandt in the collection Otterloo?", where the user only applied the location facet, and the question "Where were Rembrandt's children baptized?", where the user only filtered for the value "baptism, marriage and burial records" and found the answer within the returned primary sources.





fair amount of divergence among users, especially in preferences between faceted search or free text queries. User 3, for example, hardly used any facets, while user 2 used more facets than text queries. The average satisfaction scores of the users range between 2.1 and 3.6.

The relation between number of interactions and answer satisfaction is analysed in Figure 5. It shows that for the questions with the lowest satisfaction score, the highest number of queries was needed and for the questions with the highest satisfaction score, the fewest queries were needed. Kendall's τ between the satisfaction score and the number of queries is -1 (*p*=0.027). There is no significant relationship between the satisfaction score and the number of facets (τ =-0.2, *p*=0.81) or the number of verticals (τ = -0.4, *p*=0.46).



Figure 5. The number of facets, queries and verticals used, per answer satisfaction score

5.2.d. How do users formulate and reformulate queries (query modification patterns)?

The query modification patterns in the RemBench data were analysed and compared to patterns found for web search [Hinne, van der Heijden, Verberne, and Kraaij 2011]. The following patterns of query reformulation were distinguished [Lau and Horvitz 1999]:

- Request for additional results: the current query is the same as the previous one (same query is reissued, likely with different facet values);
- Generalization: the current query is a substring of the previous query. E.g. the query "students Rembrandt" followed by the query "students";
- Specialization: the previous query is a substring of the current query. E.g. the query "winter" followed by the query "winter landscape";
- Reformulation: the current query has at least one word in common with the previous query, but one is not a substring of the other. E.g. the query "Rembrandt England" followed by the query "Rembrandt travels";
- New topic: the current query has no words in common with the previous query.^[13] Note that the first query issued by a user by definition starts a new topic.

The occurrences of these patterns in the queries in the data were counted. The results are in Table 4. The results show that just like in web search, the most frequent pattern is "new topic". This is not surprising since every first query of a session starts a new topic. The proportion of "new topic", however, is lower than in web search data, because RemBench sessions contain more queries on average – hence a smaller portion of queries are the start of a new topic. An important difference with web search data is that generalization, specialization and reformulation are similarly

48

frequent in the RemBench data, whereas the generalization and specialization are much more infrequent in web search data. This indicates that it was relatively difficult for the users to find the precise results they were looking for (not too specific, not too general). In addition, for some topics, the reformulations are shifts from Dutch to English or from English to Dutch, e.g. the query "Rembrandt stadhuis" followed by the query "Rembrandt city hall".

	In the RemBench data		In web search data
Number of queries	150		2000000
New topic	53	35.3%	67.0%
Generalization	26	17.3%	2.2%
Reformulation	25	16.7%	12.4%
Specialization	24	16.0%	4.9%
New topic in same session	16	10.7%	16.6%
Request for additional results	6	4.0%	13.5%

 Table 4. Distribution of query modification patterns in the RemBench data compared to the distribution found for web search by Hinne et al. [Hinne, van der Heijden, Verberne, and Kraaij 2011]

5.3. User satisfaction

5.3.a. How does usability satisfaction relate to answer satisfaction?

The participants answered the questions "How satisfied are you with the answer(s) found?" (*answer satisfaction*) and "How satisfied are you with the use of RemBench for answering the question?" (*usability satisfaction*). Scores were given on a rating scale of 1-5, 5 being the highest satisfaction score.

- The mean score that was obtained for answer satisfaction is 2.90, with a standard deviation of 1.46;
- The mean score that was obtained for usability satisfaction is 2.84, with a standard deviation of 1.27.

We studied the 15 questions that were addressed by two participants, in order to get an impression of the agreement between the participants. This agreement was low: weighted Cohen's κ = 0.06, which indicates that the agreement was only slightly higher than chance agreement. This suggests that the evaluative judgments heavily depend on the individual users.

We found a strong positive relationship between answer satisfaction and usability satisfaction (Pearson's r=0.91, N=54). ^[14] This indicates that usability satisfaction was dependent on answer satisfaction: If the user was not able to find the answer with RemBench, then both the satisfaction with the answer and with RemBench are likely to be low. This might also be one of the causes of the low agreement between two users on the same question: a user who was able to find the answer will be more positive than a user who did not find the answer.

5.3.b. How does user satisfaction differ between question types?

From the satisfaction scores per question, it can be deduced which questions are more difficult to find answers to (low satisfaction scores), and which are easier (high satisfaction scores). The lists of most difficult and easiest questions are shown in Table 5. The table confirms that the easier questions are generally closed questions and factoid questions starting with "when", "where", "how long", etc. Many of the difficult questions are the list questions, which require exhaustive searching ("find all ...", "how many ..."). Figure 6 shows the average answer satisfaction score and average usability satisfaction score for the three question types in our data: closed questions, factoid questions, and list questions. According to the Mann-Whitney U-Test (for ordinal data in independent samples), the difference in answer satisfaction between closed questions and list questions is significant at the 0.05-level (p= 0.02).

49

50

51

Difficult questions (Answer satisfaction score is 1)	Easy questions (Answer satisfaction score is 5)
What is the biggest (and smallest) painting by Rembrandt?	Was Rembrandt's <i>Reading Woman</i> in the Rijksmuseum painted on canvas or panel?
How many brothers does Saskia have and when were they born?	How long has Willem Drost lived in Italy?
Are there any male portraits by Rembrandt known that are similar size to his Portrait of a young woman with a dog in Toledo?	Where were Rembrandt's children baptized (in which church)?
Find mezzotints after Rembrandt's works	When was Aert de Gelder born?
Where is Rembrandt's Sacrifice of Isaac?	What is the provenance of the <i>Night Watch</i> ^{?[15]}
Did Rembrandt have grandchildren; if yes, where were they born?	Is Ferdinand Bol older or younger than Govert Flinck?
Find literature about Rembrandt by Haverkamp-Begemann.	
Find all works by Rembrandt with a sword on them	
Find still lifes by Rembrandt and literature about them	

Table 5. Analysis of results, showing the questions that are the most difficult and the easiest to answer, according to their satisfaction scores. The questions have been translated from Dutch to English.



Figure 6. The average answer satisfaction score and average usability satisfaction score for the three question types in our data

5.3.c. What positive and negative comments did the users make?

All participants wrote down positive and negative aspects of RemBench. The eighth and ninth participant did not bring in any new aspects, which confirms that there were sufficient users to reach saturation in the reported usability issues. The lists of positive aspects and the lists of negative aspects provided by the participants in the post-task questionnaire were merged, and sorted by topic. The results are in Table 6.

 Table 6 shows that the users were predominantly positive about the graphical user interface, the interaction design, and
 54

 the content of the underlying databases. One exception was a critical note on the quality of the images, a problem that
 54

was also mentioned by Beaudoin and Brady [Beaudoin and Brady 2011] as an important factor for art historians when they use digital image libraries. The quality of the images is constrained by the copyrights of the image owners, and cannot be influenced by developers of a federated search engine that retrieves the images.

Table 6 also shows that the users were most critical about the search functionalities. Below, we discuss the numbered comments in the rightmost column of Table 6.

Торіс	Positive	Negative
Graphical user interface	-Clear colours, good font type, nice overview, good layout	-
Interaction design	-Easy to use -Search options are immediately clear -Results ordered by type	-Sometimes difficult to choose between many search options
Search functionality	-Searching possible in English and Dutch -Good filter options (especially date range and location)	 -It is not possible to search for works by one specific artist -Differences in results between English and Dutch queries -It is difficult to find works of art with specific topics -It is sometimes difficult to find specific answers -It is difficult to get the intended number of results: either too many or too few -Searching sometimes takes a lot of time
Content	-Works of art, primary and secondary sources combined -Lots of information -Reliable sources in the search results -References to RKDexplore with even more information	-Low quality of images

 Table 6. Evaluative comments provided by the participants in the post-task questionnaire

In order to meet comment 1, functionality was added for searching works by one specific artist by adding a facet "Author/artist name" in the final version of RemBench. The inconsistency between results for English and Dutch queries (comment 2) confirms our observation in Section 5.2.d that query reformulations are sometimes translations from English from Dutch or from Dutch to English. The inconsistencies in results between the two languages are not easy to solve: The databases contain textual information in both English and Dutch, but not all information is available in both languages. For example, RemDoc contains transcriptions of 17th-century Dutch documents together with translations of the documents in modern English, while the modern Dutch translation is not included. Although this is a relevant problem for multi-lingual databases, solving it is beyond the scope of this research, because it requires expansion of the content of the (external) databases. The same holds for comment 3, about finding works of art with specific topics, because indexing depends on the metadata that is available for works of art in RKDartists (see also the work by Yee et al. [Yee, Swearingen, Li, and Hearst 2003], discussed in Section 2.2). The selection of metadata values for topics in works of art leads to a bias towards the retrieval of relevant works: only the topics that exist as labels in the metadata, and only the works that have been labelled with these topics can in theory be found with a text-based retrieval engine. For example, a painting with a dog that has not been labelled with the topic "dog", cannot be found by a text-based retrieval engine.

Since only topics of works of art that have been described in the metadata can be found, it might be valuable to expand the topical annotation of works of art in future work, for example by crowd sourcing [Trant 2009]. The value of topic annotations for image search in the historical domain has also been pointed out by Choi and Rasmussen [Choi and Rasmussen 2003]. They find that subject descriptors that represent the image content were very important for user satisfaction. Thus they recommend: "For optimal access, representations for images should provide access to the topic

56

57

of the image and objects within the image, namely, what is portrayed in the image as well as what the image is about" [Choi and Rasmussen 2003, 508].

Comments numbered 4 - 6 are related to each other. The users often had problems finding the exact answer because they got too many irrelevant results. This resulted in more time needed for finding the answer. The large amount of results was in many cases caused by the presence of the word "Rembrandt" in the query (e.g. the query "Rembrandt dogs" for the question "Did Rembrandt paint dogs?"). For a query with both terms, the search system returns all documents that contain both terms *or* one of the two terms. The Boolean operator AND could be helpful here (which is allowed in the Solr query interface, but not used by the participants), requiring that the returned results contain both terms. In the final version of RemBench, discounting of the term weight for the query term "Rembrandt" was implemented, in order to ensure that for queries with the word "Rembrandt" and some other word(s), results that contain only the other words (e.g. "dog") are ranked higher than results that contain only the word Rembrandt.^[16]

58

59

6. Discussion

In this section, the two research questions are addressed, and the implications are discussed.

6.1. Requirements for a search engine in the art-history domain

In order to answer the first research question, the questions asked by Rembrandt researchers, the interactions of the subjects with RemBench and the user satisfaction were analysed (See Section 5), leading to the following conclusions:

- RemBench users enter more queries per session than web search users; the distribution of queries per session is more similar to sessions in digital library search. This reflects the informational nature of the questions in a specific topic domain, as opposed to the high proportion of navigational queries on the web. Still, the average number of queries needed per question is relatively low (2.0), which indicates that the RemBench search interface allows for efficient information seeking.
- 2. The combination of different source types is highly important in the art history domain. Four databases were combined as four different verticals in the search interface. This was preferred by domain experts, and also highly valued by the participants of the user study. Especially the combination of works of art, primary and secondary sources was considered an important asset of RemBench.
- 3. Both free-text search and facet filtering were actively used by the observed participants. There were individual differences in the frequency of use of queries and facets: Some users preferred free-text searching, while others had a preference for filtering facets. This finding confirms the value of faceted search interfaces for domain-specific information-seeking tasks.
- 4. The users appreciated the presence of multi-lingual data (Dutch and English) in the connected databases. However, they were critical about the cross-lingual access to the data. In many cases, they had to try their queries in both English and Dutch in order to get the results that they wanted. This illustrates the challenge of managing databases with multilingual content and shows the importance of cross-lingual information access to multilingual data.
- 5. A search environment for a highly specific domain (such as Rembrandt) requires domain-dependent treatment of the data and in some cases also of the queries. Domain-specificity requires pre-selection of relevant data from the connected databases (see Section 3). In addition, term discounting for the specific query term "Rembrandt" had to be implemented, because it was difficult for the participants to get the exact results they wanted for queries containing this high frequent term (see Section 5.3.c). More analysis in future work is required in order to answer the question that arises from the analysis: under what circumstances does a highly specific topic domain require domain-specific re-weighting of query terms, and when are the ordinary term weighting variables (term frequency, inverted document frequency) sufficient?

6.2. Recommendations for the design of a user observation study

Previous research has shown that for most usability tests, the proportion of additional usability problems found when 61

adding test users quickly decreases after five users. This was confirmed by the user study: Nine participants were enough to discover all usability issues. The thinking-aloud protocol was successful. When a researcher makes extensive notes, no microphone recordings are necessary. In addition, screen capture was more useful than camera recordings, which provided very little additional information about the behaviour of the user.

There is a strong correlation between answer satisfaction and usability satisfaction, and the agreement between two users trying to find the answer to the same question is low. This confirms that information seeking tasks cannot be evaluated independently of usability. This is not necessarily problematic, but should be taken into account in the design of future usability studies for information seeking: The more difficult the task, the lower the users' judgements of the usability. List questions receive the lowest satisfaction scores (significantly lower than closed questions), and require the most interactions (significantly more than factoid questions).

62

65

66

Three limitations of the presented user study should be acknowledged: First, the introduction of an externally imposed information need (a question formulated by someone else) is an efficient way of studying information-seeking behaviour. It implies, however, that the participants were not the "owners" of the questions. As a result, no strong conclusions can be drawn about user effort (number of queries issued, number of results inspected), because the amount of effort that a user puts into a search task depends on the urgency and importance of his research question – aspects that are unknown for non-owners. In addition, research by Chouldechova and Mease [Chouldechova and Mease 2013] has shown that topic owners are better capable of judging the relevance of retrieved results than non-owners. Especially for the questions that required an (exhaustive) search for elements of a list (e.g. "list all...", "how many...") it was difficult for a user who is not the owner of the information need to judge the completeness of the answer.

The second limitation is that the retrieval effectiveness as perceived by the user is limited by the content of the underlying databases. Information that is not contained in the database, such as topical metadata labels for images, cannot be retrieved by the user. In addition to that, users tend to trust the information retrieved and the ranking by the search engine [Pan, Hembrooke, Joachims, Lorigo, Gay, and Granka 2007]: they click on the highest-ranked results. If users have high trust with the search engine, while relevant information is missing from the underlying databases or the relevant information is ranked too low by the search engine, the results found by the users for their research questions are biased [Keane, O'Brien, and Smyth 2008]. This effect is stronger if users do not know what they have *not* found. We saw that large individual differences exist in how persevering and critical users are when they try to find the information they need. In future research, this user characteristic should be taken into account: when does a user decide to stop searching, either because he (thinks he) has found the information he needed, or because he gives up on finding the information [Baskaya, Keskustalo, and Järvelin 2013]?

The third limitation of the study is that all participants were students. One risk of this is that the information-seeking behaviour of students differs from the behaviour of (older) researchers [Weiler 2005]. According to Rowlands et al. [Rowlands, Nicholas, Williams, Huntington, Fieldhouse, Gunter, ... and Tenopir 2008], the generation who grew up with Google-style search relies "heavily on search engines, view rather than read and do not possess the critical and analytical skills to assess the information that they find on the web" [Rowlands, Nicholas, Williams, Huntington, Fieldhouse, Gunter, ... and Tenopir 2008, 290]. On the other hand, the authors state that the impact of ICT on this generation should not be overestimated, and that "we are all the Google generation, the young and old, the professor and the student and the teacher and the child." In addition, university students may act more as researchers than as ad hoc searchers when they are put in a working context such as the RemBench user study. Future research should address the differences between students and researchers in their information-seeking behaviour.

7. Conclusion

A user observation study with RemBench, a domain-specific federated search engine in the art history domain, provided several important insights. First, the study confirmed the importance of different source types (primary and secondary sources, textual and graphical sources) for the art history domain, which is not an atypical domain in the humanities; e.g. archaeology, media studies, historical literary and linguistic studies often work with combinations of textual and graphical sources. We showed that different source types from multiple databases can successfully be combined in a

federated search engine. Second, we found that although there are large individual differences in the preferred information seeking style, *faceted search* is a highly valued functionality. In addition, the study presented in this paper showed that the usability of domain-specific search engines can be evaluated successfully using a thinking-aloud protocol with a small number of participants, but that the measures usability satisfaction depends on the difficulty of the task given to the user.

Acknowledgements

The research for this paper was funded by CLARIN-NL under grant number CLARIN-NL-12-022. We thank our partners Huygens ING, the Netherlands Institute for Art History RKD and our colleagues from the art history department for their collaboration.

Notes

- [1] http://www.remdoc.org
- [2] http://rembench.huygens.knaw.nl/
- [3] https://rkd.nl/en/
- [4] http://www.remdoc.org
- [5] http://www.ru.nl/library/vm/ruquest
- [6] http://lucene.apache.org/solr/
- [7] BB FlashBack Express, http://www.bbsoftware.co.uk/BBFlashBack_FreePlayer.aspx
- [8] These interpretations were expressed by the users because of the thinking-aloud set-up
- [9] A permalink is a unique URL that points to one specific record in a database. A permalink is meant to be stable for a long period of time. For example, the permalink to the entry "the Nightwatch" in RKDimages is https://rkd.nl/en/explore/images/13893
- [10] A search session is defined as the series of all actions related to one question [Sriram, Shen, and Zhai 2004].
- [11] Subsequent duplicate queries were removed from the TEL sessions, as explained in Verberne et al. 2010 [Verberne, Hinne, van der Heijden, Hoenkamp, Kraaij, van der Weide 2010]
- [12] Workshop on Web Search Click Data: http://research.microsoft.com/en-us/um/people/nickcr/wscd09/
- [13] Note that this is a pragmatic, not a formal, definition. Two subsequent queries can still be on the same topic if they have no words in common.
- [14] If the 5-point judgements would not be considered interval data but ordinal data, then Kendall's τ should be used. This test also indicates a strong positive relationship between the two variables (Kendall's τ = 0.85, p<0.001).
- [15] This is in theory a question that might require exhaustive searching (find all former locations of the Night Watch), but since the complete provenance of the painting is listed in one record RKDimages, it was easily found by the user
- [16] It should be noted here that if a search system provides good ranking of the results, the Boolean AND operator should not be necessary, because results with both terms present should be ranked higher in the result list than results with one of the two terms. Also, query terms that occur in few documents ("dogs") should be weighted heavier than query terms that occur in many documents ("Rembrandt"). Although this term characteristic ("inverted document frequency") is a component of the ranking algorithm in Solr (see http://www.solrtutorial.com/solr-search-relevancy.html), it was sometimes difficult for the participants to get the results they wanted for queries with one or more high frequent terms.

Works Cited

Baskaya, Keskustalo, and Järvelin 2013 Baskaya, F., Keskustalo, H., & Järvelin, K. (2013, October). "Modeling behavioral factors ininteractive information retrieval". In *Proceedings of the 22nd ACM international conference on*

Conference on information & knowledge management (pp. 2297-2302). ACM.

- Bates 1996a Bates, M. J. (1996). "Document familiarity, relevance, and bradford's law: The Getty online searching project report no. 5". *Information Processing & Management*, 32(6), 697-707.
- Bates 1996b Bates, M. J. (1996). "The Getty end-user online searching project in the humanities: Report no. 6: Overview and conclusions". *College & Research Libraries*, 57(6), 514-523.
- Bates, Wilde, and Siegfried 1993 Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). "An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report Number 1". *The Library Quarterly*, 1-39.
- Bates, Wilde, and Siegfried 1994 Bates, M. J. (1994). "The design of databases and other information resources for humanities scholars: The Getty Online Searching Project Report No. 4". Online and CD-ROM review, 18(6), 331-340.
- Bates, Wilde, and Siegfried 1996 Bates, M. J., Wilde, D. N., & Siegfried, S. (1996). "Research practices of humanities scholars in an online environment: The Getty online searching project report no. 3". *Library & Information Science Research*, 17(1), 5-40.
- **Beaudoin and Brady 2011** Beaudoin, J. E., & Brady, J. E. (2011). "Finding visual information: A study of image resources used by archaeologists, architects, art historians, and artists". *Art Documentation: Journal of the Art Libraries Society of North America*, 24-36.
- Choi and Rasmussen 2003 Choi, Y., & Rasmussen, E. M. (2003). "Searching for images: the analysis of users' queries for image retrieval in American history". *Journal of the American society for information science and technology*, 54 (6), 498-511.
- Chouldechova and Mease 2013 Chouldechova, A., & Mease, D. (2013). "Differences in search engine evaluations between query owners and non-owners". In Proceedings of the sixth ACM international conference on Web search and data mining (pp. 103-112). ACM.
- **Davies 2007** Davies, K. (2007). "The information-seeking behaviour of doctors: a review of the evidence". *Health Information & Libraries Journal*, 24(2), 78-94.
- Dudek, Mastora, and Landoni 2007 Dudek, D., Mastora, A., & Landoni, M. (2007). "Is Google the answer? A study into usability of search engines". *Library Review*, 56(3), 224-233.
- George, Bright, Hurlbert, Linke, St Clair, and Stein 2006 George, C. A., Bright, A., Hurlbert, T., Linke, E. C., St Clair, G., & Stein, J. (2006). "Scholarly use of information: graduate students' information seeking behaviour". *Information Research*, 11(4).
- Gerjets, Kammerer, and Werner 2011 Gerjets, P., Kammerer, Y., & Werner, B. (2011). "Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data". *Learning and Instruction*, 21(2), 220-231.
- **Gossen, Hempel, and Nürnberger 2013** Gossen, T., Hempel, J., & Nürnberger, A. (2013). "Find it if you can: usability case study of search engines for young users". *Personal and ubiquitous computing*, 17(8), 1593-1603.
- Granka, Joachims, and Gay 2004 Granka, L. A., Joachims, T., & Gay, G. (2004, July). "Eye-tracking analysis of user behavior in WWW search". In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 478-479). ACM.
- Hemminger, Lu, Vaughan, and Adams 2007 Hemminger, B. M., Lu, D., Vaughan, K. T. L., & Adams, S. J. (2007). "Information seeking behavior of academic scientists". *Journal of the American Society for Information Science and Technology*, 58(14), 2205-2225.
- Hermjakob, Hovy, and Lin 2000 Hermjakob, U., Hovy, E. H., & Lin, C. Y. (2000). "Knowledge-based question answering". In Proceedings of the Sixth World Multiconference on Systems, Cybernetics, and Informatics (SCI-2002).
- Hinne, van der Heijden, Verberne, and Kraaij 2011 Hinne, M., van der Heijden, M., Verberne, S., & Kraaij, W. (2011). "A multi-dimensional model for search intent". In *Proceedings of the Dutch-Belgium Information Retrieval workshop* (pp. 20-24)
- Jansen and Spink 2006 Jansen, B. J., & Spink, A. (2006). "How are we searching the World Wide Web? A comparison of nine search engine transaction logs". *Information Processing & Management*, 42(1), 248-263.
- Jansen, Spink, and Saracevic 2000 Jansen, B. J., Spink, A., & Saracevic, T. (2000). "Real life, real users, and real needs: a study and analysis of user queries on the web". *Information processing & management*, 36(2), 207-227.

Jascó 2004 Jacsó, Péter (2004). "Thoughts about federated searching". Information Today: 21(9).

- Keane, O'Brien, and Smyth 2008 Keane, M. T., O'Brien, M., & Smyth, B. (2008). "Are people biased in their use of search engines?". Communications of the ACM, 51(2), 49-52.
- Kules, Capra, Banta, and Sierra 2009 Kules, B., Capra, R., Banta, M., & Sierra, T. (2009, June). "What do exploratory searchers look at in a faceted search interface?". In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 313-322). ACM.
- Lalmas 2011 Lalmas, M. (2011). "Aggregated search". In Advanced Topics in Information Retrieval (pp. 109-123). Springer Berlin Heidelberg.
- Lau and Horvitz 1999 Lau, T., & Horvitz, E. (1999). Patterns of search: analyzing and modeling web query refinement (pp. 119-128). Springer Vienna.
- Makri, Blandford, and Cox 2008 Makri, S., Blandford, A., & Cox, A. L. (2008). "Investigating the information-seeking behaviour of academic lawyers: From Ellis's model to design". *Information Processing & Management* 44 (2), 613-634.
- **Meho and Tibbo 2003** Meho, L. I., & Tibbo, H. R. (2003). "Modeling the information-seeking behavior of social scientists: Ellis's study revisited". *Journal of the American society for Information Science and Technology* 54(6), 570-587.
- Nielsen and Landauer 1993 Nielsen, J., & Landauer, T. K. (1993, May). "A mathematical model of the finding of usability problems". In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 206-213). ACM.
- Pan, Hembrooke, Joachims, Lorigo, Gay, and Granka 2007 Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). "In google we trust: Users' decisions on rank, position, and relevance". *Journal of Computer-Mediated Communication*, 12(3), 801-823.
- **Ponnuswami, Pattabiraman, Wu, Gilad-Bachrach, and Kanungo 2011** Ponnuswami, A. K., Pattabiraman, K., Wu, Q., Gilad-Bachrach, R., & Kanungo, T. (2011, February). "On composition of a federated web search result page: using online users to provide pairwise preference for heterogeneous verticals". In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 715-724). ACM.
- Rieh 2006 Rieh, S. Y. (2006). "Analysis of multiple query reformulations on the web: The interactive information retrieval context". *Information Processing & Management*, 42 (3), 751-768.
- **Rodríguez Ortega 2013** Rodríguez Ortega, N. (2013). "It's Time to Rethink and Expand Art History for the Digital Age". *The Getty Iris. Retrieved from* http://blogs.getty.edu/iris/its-time-to-rethink-and-expand-art-history-for-the-digital-age
- Roscam Abbing 2006 Roscam Abbing, M. (2006). Rembrandt 2006: the new Rembrandt Documents. Leiden.
- Rowlands, Nicholas, Williams, Huntington, Fieldhouse, Gunter, ... and Tenopir 2008 Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., ... & Tenopir, C. (2008, July). "The Google generation: the information behaviour of the researcher of the future". In *Aslib Proceedings* (Vol. 60, No. 4, pp. 290-310). Emerald Group Publishing Limited.
- Russell-Rose, Lamantia, and Makri 2014 Russell-Rose, T., Lamantia, J., & Makri, S. (2014). "Defining and Applying a Language for Discovery". In *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation* (pp. 3-28). Springer International Publishing.
- Sadeh 2007 Sadeh, T. (2007). "Time for a change: new approaches for a new generation of library users". *New Library World*, 108 (7/8), 307-316.
- Shokouhi and Si 2011 Shokouhi, M., & Si, L. (2011). "Federated search". *Foundations and Trends in Information Retrieval*, 5(1), 1-102.
- Siegfried, Bates, and Wilde 1993 Siegfried, S., Bates, M. J., & Wilde, D. N. (1993). "A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project Report No. 2". Journal of the American Society for Information Science, 44(5), 273-291.
- **Spink 2002** Spink, A. (2002). "A user-centered approach to evaluating human interaction with web search engines: an exploratory study". *Information processing & management*, (3), 401-426.
- Spink and Jansen 2006 Spink, A., & Jansen, B. J. (2006). "Searching multimedia federated content web collections". Online Information Review, 30(5), 485-495.

Sriram, Shen, and Zhai 2004 Sriram, S., Shen, X., & Zhai, C. (2004, July). "A session-based search engine". In

Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 492-493). ACM.

Strauss and van der Meulen 1979 Strauss, W.L. and van der Meulen, M. (1979) The Rembrandt documents. New York.

- **Tenopir, King, Spencer, and Wu 2009** Tenopir, C., King, D. W., Spencer, J., & Wu, L. (2009). "Variations in article seeking and reading patterns of academics: What makes a difference?". *Library & Information Science Research*, (3), 139-148.
- **Trant 2009** Trant, J. (2009). "Tagging, folksonomy and art museums: Early experiments and ongoing research". *Journal of Digital Information*.
- **Tunkelang 2009** Tunkelang, D. (2009). "Faceted search". Synthesis lectures on information concepts, retrieval, and services, 1 (1), 1-80.
- **Van Waes 2000** Van Waes, L. (2000). "Thinking aloud as a method for testing the usability of websites: the influence of task variation on the evaluation of hypertext". *Professional Communication, IEEE Transactions on*, 43(3), 279-291.
- Verberne, Hinne, van der Heijden, Hoenkamp, Kraaij, van der Weide 2010 Verberne, S., Hinne, M., van der Heijden, M., Hoenkamp, E., Kraaij, W., van der Weide, T.P. (2010). "How does the library searcher behave? A contrastive study of library search against ad-hoc search". In: Braschler, M., Harman, D., Pianta, E. (eds.): *CLEF 2010 Labs and Workshops*. Abstracts Notebook Papers, CLEF 2010, p. 99
- Vezzosi 2009 Vezzosi, M. (2009). "Doctoral students' information behaviour: an exploratory study at the University of Parma" (Italy). *New Library World*, 110(1/2), 65-80.

Voorhees 2003 Voorhees, E. M. (2003). "Overview of TREC 2003". In TREC (pp. 1-13).

- Weiler 2005 Weiler, A. (2005). "Information-seeking behavior in Generation Y students: Motivation, critical thinking, and learning theory". *The Journal of Academic Librarianship*, 31(1), 46-53.
- White and Drucker 2007 White, R. W., & Drucker, S. M. (2007, May). "Investigating behavioral variability in web search". In Proceedings of the 16th international conference on World Wide Web (pp. 21-30). ACM.
- Yee, Swearingen, Li, and Hearst 2003 Yee, K. P., Swearingen, K., Li, K., & Hearst, M. (2003, April). "Faceted metadata for image search and browsing". In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 401-408). ACM.
- **Zorich 2012** Zorich, D. (2012). "Transitioning to a Digital World: Art History, Its Research Centers, and Digital Scholarship". *Report to the Samuel H. Kress Foundation and the Roy Rosenzweig Center for History and New Media, George Mason University.* Retrieved from http://journalofdigitalhumanities.org/1-2/transitioning-to-a-digital-world-by-diane-zorich/



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.