

RemBench: A Digital Workbench for Rembrandt Research

Suzan Verberne^{a,d}, Rudie van Leeuwen^b, Guido Gerritsen^c and Lou Boves^a

^aCentre for Language Studies, Radboud University

^bCentre for Historical, Literary and Cultural Studies, Radboud University

^cHuygens Institute for the History of the Netherlands

^dLeiden Institute for Advanced Computer Science, Leiden University

Corresponding author: s.verberne@liacs.leidenuniv.nl

ABSTRACT

In this chapter, we present RemBench, a search engine for research into the life and works of Rembrandt van Rijn. RemBench combines the data from four different databases behind one interface using federated search technology. Metadata filtering is enabled through faceted search. RemBench enables art historians and other professionals interested in Rembrandt's time to find all the information on Rembrandt that is available in online repositories through one application. The functionality and user interface of RemBench were developed in close collaboration with domain experts, and evaluated in a user study with nine students of history and art history. We found that the users were positive about the usability of RemBench, especially about its user interface and interaction design. We think that RemBench sets an example for search engines in the digital humanities. Our most important recommendation is the use of federated search (with different types of results in different verticals) and faceted search in the art history domain. In addition, we recommend evaluation through a user observation study, which is already possible with a small number of participants.

28.1 Introduction

Art historians typically study different types of sources: works of art, primary and secondary sources describing these works of art, the life of the artist, the provenance of the work, and the social and economic context in which the artist worked. The relevant information can be found in independent digital (and, non-digital) resources, developed by museums, archives and libraries.

How to cite this book chapter:

Verberne, S, van Leeuwen, R, Gerritsen, G and Boves, L. 2017. RemBench: A Digital Workbench for Rembrandt Research. In: Odijk, J and van Hessen, A. (eds.) *CLARIN in the Low Countries*, Pp. 337–350. London: Ubiquity Press. DOI:<https://doi.org/10.5334/bbi.28>. License: CC-BY 4.0

More and more sources are being digitised and made accessible online (Rodríguez Ortega, 2013). However, with each source type residing in its own database, art historians have to use a number of independent digital applications in parallel, each with their idiosyncratic interface. In the worst case, their research even requires visiting several museums, libraries or archives in different cities.

In this chapter we describe the development and evaluation of a specialised search engine in the art history domain that integrates multiple databases behind one interface: RemBench.¹ RemBench stands for ‘a Digital Workbench for Rembrandt Research’. The overall goals of the RemBench project were (1) to develop a working environment in which researchers have easy access to all the information about the 17th-century Dutch artist Rembrandt van Rijn and his environment and (2) to show the value of online digital data for art history research.

RemBench was developed in the context of CLARIN-NL, a large national project in the Netherlands (2009–2015) which aimed to improve the research infrastructure for humanities researchers who work with language data and tools. Data and tools developed in the context of CLARIN-NL are available to outside researchers via the CLARIN infrastructure, thereby making humanities research that requires language resources ‘easier, faster, better, and in some cases even possible for the first time.’² Within CLARIN-NL, RemBench was the only project directed at the art history domain. The art history domain is an interesting case for search engines tailored to digital humanities, because relevant information is typically contained in diverse sources (primary sources as well as secondary ones), and the data can be in textual as well as in graphical form.

RemBench is a *faceted search* engine: free text search is combined with filtering functionality for metadata values (Tunkelang, 2009). Access to four different databases is realised through *federated search*. Federated search (sometimes called distributed information retrieval) is a technique for searching multiple collections simultaneously for one single query; the results returned by selected collections are integrated into a single result page (Jacsó, 2004; Shokouhi and Si, 2011). In the context of CLARIN, a federated search infrastructure was developed for enabling the search for suitable language resources (Stehouwer et al., 2012): researchers can use the infrastructure to search through the content of multiple language resources (corpora) that potentially contain useful information for answering their research question. This type of federated search makes high demands on the unification of metadata.

Art historians who study the life and works of Rembrandt van Rijn are the primary target group of RemBench. We identified several additional target groups, such as historians with questions related to 16th–18th century life, literary scholars, linguists studying 17th-century Dutch, genealogists, and economists studying the art market in the Golden Age.

The contributions of this project are threefold: (1) the integration of the metadata of four Rembrandt-related databases; (2) the development of a user-friendly search engine for Rembrandt research that gives access to distributed resources; and (3) recommendations for the evaluation of search engines in the art history domain.

28.2 About RemBench

RemBench connects four existing databases. The first two are RKDartists and RKDimages,³ two art historical databases maintained by the Netherlands Institute for Art History (RKD). 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.

¹ <http://rembench.huygens.knaw.nl/>

² <http://www.clarin.nl/>

³ <http://explore.rkd.nl/>

The third database connected to RemBench is RemDoc,⁴ a digital collection of primary documents that relate to the life and works of Rembrandt van Rijn. In the RemDoc project, all known documents that relate to Rembrandt, as a person and as an artist, as well as to his ancestors, family, and business partners, from the 15th to the 18th century have been collected and published. The database contains 1,667 documents.

The fourth database in RemBench is a university library catalogue, digitised and made available through the RUQuest⁵ system, a library search system that provides access to the full collection of the Radboud University Library and to the full-text articles of all journals that Radboud University has subscribed to.

The project consisted of two phases: first, the metadata of the four different 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 RemBench user interface was designed by a professional designer, in interaction with art historians, who added their specific wishes.

The architecture of RemBench is shown in Figure 28.1; its user interface in Figure 28.2. The URL of RemBench is <http://rembench.huylgens.knaw.nl/>.

28.3 Challenges

In this section, we summarise the challenges that we faced in the RemBench project: (1) selecting the relevant data and metadata fields from four different databases; (2) the interoperability of the resources; (3) developing the search engine; and (4) designing the user interface. The following four subsections address these challenges.

28.3.1 Selecting Data and Metadata Fields

For the integration of the four databases, the following three questions needed to be answered: (a) which subset of the data from each database should be made searchable? (b) which metadata fields should be shown as facets in the faceted search? and (c) which fields should be searched with free text search?

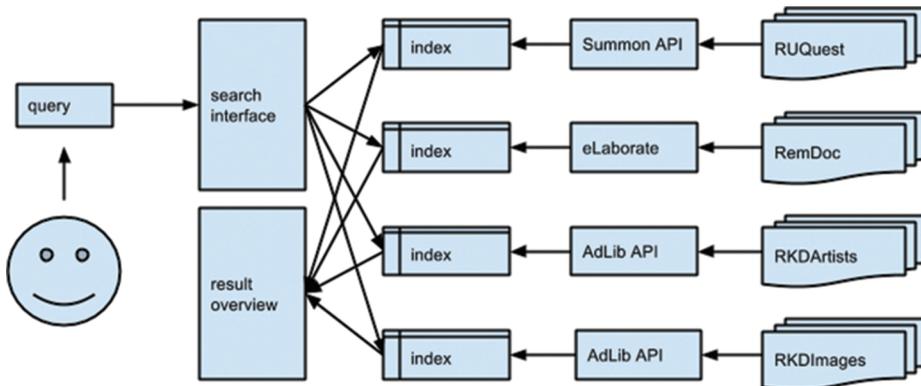


Figure 28.1: The architecture of RemBench

⁴ <http://www.remdoc.org/>

⁵ <https://ru.on.worldcat.org/discovery>

Figure 28.2: The user interface of RemBench

28.3.1.1 Data Selection from the Databases

We made a selection of records from the RKD databases by including only 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 that are mentioned in the Rembrandt record (either under ‘pupil of’, ‘teacher of’, ‘followed by’, ‘influenced by’ or ‘had influence on’) were included. In the selection of images, all works of art from RKDimages that have been attributed to one of the artists in the artist selection were included. The resulting selection consists of 1,857 works of art and 59 artists.

No selection needed to be performed in RemDoc, because all content in RemDoc is related to Rembrandt.

In RUQuest, we planned to harvest all items that were returned from the collection for the search term ‘Rembrandt’ (84,081 results). This would ensure a filtering for the domain, so that RemBench queries will only give results that are relevant to the Rembrandt domain of research. (Without filtering, a query such as ‘saskia’ would give many results that are not about Rembrandt’s wife Saskia van Uylenburgh.) However, the API that gave access to RUQuest could only return a maximum of 1,000 records. Consequently, not all records relevant to Rembrandt could be harvested at index time. As an alternative to pre-fetching all 84,081 records, we retrieved the maximum of 1,000 results through the API for each query at query time, and stored the results in a local index. The local index was searched again for every following query, together with a call to the API for another 1,000 results.

28.3.1.2 *Field Selection for Faceted Search*

In the RemBench interface, free text search is combined with faceted search (filtering for metadata values, see Section 28.3.2). We initially chose three metadata fields for faceted search: Content type, Artist/author type, and Location, and included Publication date as an additional filter. Later, we added Author/Artist name, to allow filtering by one specific artist or author. For each of these fields, we mapped the values for the different databases onto each other. For example, for content type, we defined a value ‘articles in magazines/journals’, which corresponds to six different content type values in RUQuest (‘Journal article’, ‘Newsletter article’, etc.) and one value for Document type in RemDoc: ‘periodical/magazine’. Not all facets have a corresponding metadata field in all databases. For example, when the user selects a value for the Content Type facet, RKDartists does not return any results.

28.3.1.3 *Field Selection for Free Text Search*

In the original databases there is a distinction between free text fields and restricted metadata fields. In the local index, we indexed all fields with textual content as free text fields. Table 28.1 shows the fields that were indexed for free text search.

28.3.2 *Interoperability of the Resources*

After selecting the data and metadata field from the databases that are used for faceted search, we needed to map the metadata categories onto each other. The central step was to map the resource-specific metadata categories to ISOcat data categories (DCs). ISOcat distinguishes between *Complex DCs*, which are (meta)data fields that can have a value, and *Simple DCs*, which are the values themselves. For example, the RemBench Facet ‘Context type’ (a complex DC) can have as values ‘altar piece’, ‘articles in magazines/journals’, ‘baptism, marriage and burial records’ etc. (simple DCs). Table 28.2 shows an overview of the complex DCs that we defined in ISOcat. The first two, date and author, already existed as DCs with the same meaning in ISOcat.

The majority of the complex DCs are open categories: they can get any value (either a string or a date). When harmonising the contents of multiple databases, the content of these fields needed to be standardised, especially in the case of names. For example, one database might use ‘Paris’ as name, while another uses ‘Parijs’ (the Dutch word for Paris) and a third one ‘Paris, France’. For the standardisation of geographical names, we used the controlled vocabularies by Getty.⁶ Note that disciplines unrelated to art might have preferred a different vocabulary, for example the UNGEGN World Geographical Names database.⁷

⁶ <https://www.getty.edu/research/tools/vocabularies/tgn/>

⁷ <http://unstats.un.org/unsd/geoinfo/geonames/>

Database	Field	Explanation/comments	
RemDoc	Entry Name		
	Translation		
	Diplomatic	This field contains the transcription	
	Comments		
RUQuest	SubjectTerms		
	Title		
	PublicationTitle		
	Author		
RKDimages	Abstract		
	i2 'benaming_kunstwerk' i3 'andere_benaming' i4 'titel_engels'	Title of work of art, alternative title and English title	
	i8 'naam'	Artist	
	i75 'RKD_algemene_trefwoorden'	General keywords. Example value: 'oude testament & apocriefen' (old testament & apocryphals)	
	i50 'collectienaam'	Collection	
	i96 'inbrenger'	Contributor	
	i97 'naam_koper'	Buyer's name	
	i29 'opdrachtgever'	Patron	
	i123 'persoonsnummer'	Person number (portrayed person)	
	i102 'veilinghuis'	Auction house	
	i122 opmerking_algemeen	General comment	
	RKDartists	a1 'kunstenaarsnaam'	Name of artist
		a7 'spelling_variant'	Spelling variant of name
a29 'plaats_van_werkzaamheid'		Place of activity	
a34 'kwalificatie'		Qualification ('role'/profession)	
i122 opmerking_algemeen		General comment	

Table 28.1: Fields from each of the databases that were indexed for free text search

Three DCs are closed with respect to the values they can get: 'Content Type', 'Archive Main Type' and 'Person Type'. The values that the closed categories can take have been defined in the simple DCs in ISOcat. An example of a simple DC is 'articlesInMagazinesJournals', which is a possible value for the complex DC 'Content Type'. All data categories (complex and simple) that are used in a project are defined in the Data Category Selection (DCS). We delivered the RemBench-DCS to CLARIN-NL as an XML file in the Data Category Interchange Format (DCIF).

In addition, we defined the RemBench metadata on the level of resources in the Component MetaData Infrastructure (CMDI), which provides a framework to describe and reuse metadata formats. We submitted three entries to CMDI for each of the three resources contained in RemBench: RemDoc, RKDexplore and RUQuest; these entries can be found in CLARIN's Virtual Language Observatory.⁸

⁸ <https://vlo.clarin.eu/>

DC name	Field in RemDoc	Field in RKD
date (existing DC-4335)	Date	i13 datering
author (existing DC-4115)	Author	
English Translation	Translation	
Name of Object	Name of object	i2 benaming kunstwerk
Location of Object	Location	
Content Type	Document Main Type	i18 objectcategorie
Archive Main Type	Archive Main Type	
Person Type	Role	a34 kwalificatie persoon
name of artist		a1 kunstenaarsnaam=i8 naam
date of birth		a17-a18 geboortedatum_begin en geboortedatum_eind
date of death		a22-a23 sterftedatum_begin en steftedatum_eind
period of activity		a26-a27 werkzame_periode_begin en werkzame_periode_eind
place of activity		plaats_van_werkzaamheid
text transcription	Diplomatic	
secondary document	Comments	

Table 28.2: List of complex data categories defined for RemBench in ISOcat.

28.3.3 Developing the Search Engine

The effectiveness of faceted search for research in the art history domain has been shown by Yee et al. (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 found 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.

The technology for faceted⁹ 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.¹⁰ With Solr, it is possible to search in multiple fields from multiple databases at the same time (federated search). The user's query is forwarded from the RemBench interface to each of the database-specific indexes (see Figure 28.1), and its content is matched to each of the text fields listed in Table 28.1. The results are returned from each of the databases in order of relevance.

28.3.4 Designing the User Interface

The four databases were treated as four separate verticals in the interface, which did not merge 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 clearer to the user. Keeping verticals separate avoids the challenge of ranking the results from the different verticals relative to each other (Ponnuswami et al., 2011). In the user interface, the 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

⁹ <https://github.com/HuygensING/faceted-search>

¹⁰ <http://lucene.apache.org/solr/>

vertical, a maximum of five results are shown on the front page. To see more results (in an overlay screen), the user has to click on 'More' (cf. Figure 28.2).

28.4 Evaluation

In order to evaluate the beta version of RemBench and improve it for the final version, we set up a user observation study. In this section, we describe the study and the results.

28.4.1 Design, Materials and Procedure

Students of history and art history (bachelor's and master's level) from Radboud University were recruited to participate in the user study; there were nine participants (two male, seven female; median age: 20.5) in total. We expected that nine participants would be sufficient, because 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. The participants were paid a volunteer fee of €10.

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 method of usability testing is adopted in the current chapter. In addition, we combine screen capturing with thinking-aloud, as suggested by Van Waes (2000), to collect the user interactions.

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 28.4.2), one question at a time. Some of the questions required a single answer (yes/no, name, title, or place), others a list of items. The participants were asked to use all the functionalities of RemBench they needed to find the answers and to stop their search when they felt that they had tried everything they could to find the answers.

The participants were working on a Windows 7 PC with Firefox. A researcher loaded the RemBench homepage for them and gave them the list of questions to work on. User-system interaction was observed using a thinking-aloud set-up (Gerjets et al., 2011): the participants were asked to voice aloud their thinking process, what actions they took in the search process and why they took them. A researcher was sitting next to the participant and took extensive notes of what the participant did and said. Desktop activity was recorded using screen capture software.¹¹ After 45 minutes, the researcher asked the participant to finish the current question and skip the remaining ones.

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

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

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

¹¹ BB FlashBack Express: http://www.bbsoftware.co.uk/BBFlashBack_FreePlayer.aspx

28.4.2 Questions about Rembrandt

Two art historians from our research team (who are working on the topic of Rembrandt themselves) phrased a number of questions about Rembrandt that are likely to be addressed by Rembrandt researchers. They provided 61 questions. Some examples of questions are listed in Table 28.3. Each participant was assigned 10 of the 61 questions.

28.4.3 Results

Out of the 61 questions, 54 were addressed by at least one participant. The remaining seven were all 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. In this section, we report (a) the measured user satisfaction, (b) the outcomes of the post-task questionnaire and (c) which features of RemBench were used by the participants and which were not.

28.4.3.1 User 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 was 2.90, with a standard deviation of 1.46. The mean score that was obtained for usability satisfaction was 2.84, with a standard deviation of 1.27. We found a strong positive relationship between answer satisfaction and usability satisfaction (Pearson’s $r = 0.91$, $N = 54$). 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 were likely to be low.

28.4.3.2 Outcome of Post-Task Questionnaire

All participants wrote down positive and negative points about RemBench. The eighth and ninth participant did not bring up any new points, which confirms that there were enough 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.

We found that the users were predominantly positive about the graphical user interface, the interaction design, and the content of the underlying databases, and that they were most critical about the search functionalities. We made the following adaptations to the final version of RemBench in order to follow the advice given by the participants in their comments:

● 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> ?
● Was Rembrandt’s <i>Reading Woman</i> in the Rijksmuseum painted on canvas or panel?
● Did Rembrandt know any Jews?
● Did Rembrandt paint dogs?
● 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 28.3: A few examples of questions in our dataset (translated from Dutch)

- Functionality was added for searching works by one specific artist, by adding an ‘Author/artist name’ facet;
- 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 word(s), e.g. ‘dog’, are ranked higher than results that contain only the word ‘Rembrandt’.

A few other issues identified in the comments could not be solved because they require expansion of the content of the (external) databases. One example is the comment that it is sometimes difficult to find works of art with specific topics (e.g. dogs, snow). Since works of art can only be found through the topics that are included in the metadata, it might be valuable to expand the topical annotation of works of art in future work, for example through 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 (2003), who found that topic descriptors that represent the image content were very important for user satisfaction.

28.4.3.3 Which Features were Used by the Participants and Which were Not?

All functionalities were used by the participants, except for the Refresh button (Clear everything) and the advanced query options of fuzzy search and Boolean query operators. Instead of the Refresh button, the users manually cleared the fields when starting a new query. This sometimes led to mistakes, because they forgot to clear a facet value and then entered a new query, getting fewer results than they expected.

Fuzzy search is an option that allows the user to find non-exact matches of their search term. This can be very useful for finding spelling variants or in case the user is not sure of a specific spelling. With the fuzzy search option selected (via a checkbox above the search field), the search system can return primary documents containing the string *Rembrant* for the query ‘Rembrandt’ (in the 17th century, Dutch spelling was not normalised yet), or the painting *Storm on the Sea of Galilee* for the query ‘Galilee’. The students participating in the study either did not know the purpose of the fuzzy search option or overlooked the option in the interface.

Boolean query operators are useful to force a specific term to be presented in the result list. For example, to answer the question ‘Did Saskia have brothers?’ it might be profitable to require both terms ‘Saskia’ and ‘brothers’ to be present in the results, because there are many sources containing either one of the two. The Boolean query ‘Saskia AND brothers’ would accomplish this. Similarly, to answer the question ‘Did Rembrandt paint dogs?’ a user might want to require that the term ‘dog(s)’ be in any result, because there are many results returned for the term ‘Rembrandt’.¹²

In the final version of RemBench, some user guidance for these unused functionalities was added in the form of mouse-over tooltips.

28.5 Discussion

In this section, we make a number of recommendations based on the challenges that we faced in the RemBench project.

¹² 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 only one of the two terms. In addition to this, 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 highly frequent terms.

28.5.1 *Recommendations for Data Management in Digital Humanities Projects*

28.5.1.1 *Increasing the Scope of Metadata from Linguistics to Humanities*

When we started RemBench, the CLARIN-NL metadata infrastructure was fully directed at language resources with linguistic metadata. To make the CLARIN infrastructure useful for the humanities as a whole, this scope must be broadened. We made a start with RemBench, introducing the ISOcat-profile ‘Historical objects’. We were able to reuse the data categories DC-4335 date (from profile Terminology) and DC-4115 author (from profile Metadata) in ISOcat, but we had to create new data categories for the other (meta)data fields in RemBench. Interestingly, for a number of them, seemingly similar data categories existed in other profiles, such as for ‘transcription’, but these always had a different meaning. This is not a problem, and is inevitable if you create a repository across multiple domains, but in many cases it was clear from the definitions of the data categories that the creator was not aware of possible reuse outside the scope of linguistics. For example, the definition for DC-6037 translation is ‘representation in another language (of a motto)’.

In the meantime, CLARIN-EU has abandoned ISOcat in favour of the less complex CLARIN Concept Registry (CCR) and the Component MetaData Infrastructure (CMDI). This holds the promise of making the CLARIN infrastructure attractive for a broader range of humanities researchers.

28.5.1.2 *Distinction between Data and Metadata*

A clear distinction between data and metadata on the record level does not exist in all domains and for all data types. For example, if we have a collection of 18th-century hand-written documents that is stored physically in a city archive and electronically in a database, then we could consider the documents themselves, the scan of the documents, and perhaps also their transcription (computer-readable representation of the text on the scan) to be the data, and all other information (translation, comments, annotations, author, date, location) to be the metadata. But when we are indexing these fields for searching and filtering, we often consider all fields that we use for free-text search (transcription, translation, comments) as data – perhaps secondary data is a good term here. The same situation occurs for spoken language, where the written transcription can be considered as data and as metadata. This matters for the documentation of requirements for descriptive metadata, and for other locations where CLARIN-NL speaks of ‘metadata’ in documentation.

28.5.1.3 *Access to Library Data*

In the linking of library sources to RemBench, we ran into technical and IPR issues. The original plan was to use Picarta as library source. Unfortunately, there was no API for direct access to Picarta. Alternatives were WorldCat and RUQuest (implemented as Summon database by Serial Solutions). After testing both, we concluded that the WorldCat API returns very limited metadata, and is less flexible for implementations. The Summon API that gave access to the RUQuest library database was much more complete, was more flexible to work with and retrieves very extensive metadata. One disadvantage, as mentioned in section 28.3.1.1, is that the Summon API could only return a maximum of 1,000 records. The same problem may also occur in future projects, which limits the use of library data. This should be taken into account when writing project plans for future projects using library data. It also shows that CLARIN needs a facility for tracking changes in all external resources and applications that CLARIN services link to.

28.5.2 *Recommendations for the Evaluation of Search Engines in the Humanities*

Based on our experiences, we encourage researchers to conduct user observation studies to evaluate the usability of domain-specific search engines. In our user study, nine participants were enough to discover all usability issues. The thinking-aloud protocol has proven to be successful. When a researcher makes extensive notes, no microphone recordings are necessary.

There is one caveat that should be taken into account in the design of usability studies for search engines: there is a strong correlation between answer satisfaction and usability satisfaction. In other words: the more difficult the questions that the participants try to answer with the search engine, the lower their judgements of its usability.

One limitation of our 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. 2008 (2008:290), 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.' 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' (Rowlands et al., 2008). Future research should address the differences between students and researchers in their information-seeking behaviour.

28.6 Conclusions

In the RemBench project, we integrated the data and metadata from four different databases behind one search interface, to facilitate online research on the topic of Rembrandt van Rijn, his works and his relatives. The usability of RemBench was evaluated by nine users. We found that the users were positive about the usability of RemBench, especially its user interface and interaction design. They were moderately satisfied with the use of RemBench for answering Rembrandt-related questions. It appeared to be possible to develop a workable combination of faceted and federated search with an acceptable amount of effort needed for mapping and standardising metadata values.

In its current version, RemBench serves as a portal for further research. Its main value for scholars is that multiple sources are brought together at one location ('as a single bookmark in the browser', according to an art historian who studies iconography by Rembrandt; personal communication). This allows the user to immediately find the secondary literature that relates to the works of art that are the focus of study. In addition, RemBench serves an exploratory purpose: using topical queries, the user can quickly see whether a specific topic (or iconographical subject) is frequently addressed in works of art and primary documents such as inventories. If it is, the topic is potentially relevant; this a starting point for further research.

From the (art) history perspective there are three directions in which RemBench can (and should) be developed: (a) extending and integrating the metadata content of RemBench with other resources, such as the Montias Database¹³ and ECARTICO¹⁴; (b) the integration of data that provide insight into (business) relationships in the 17th century, including those of Rembrandt's students and followers in the Amsterdam art community; and (c) the development of a publication platform for newly written commentaries and other secondary literature. Access to new book and journal publications is handled via the live interface with RUQuest, but that is not the case for new commentaries.

¹³ <http://research.frick.org/montias/>

¹⁴ <http://www.vondel.humanities.uva.nl/ecartico/>

We think that RemBench sets an example for search engines in the digital humanities. Our most important recommendation is the use of federated search and faceted search in the art history domain: with federated search, it is possible to search multiple databases at once, while faceted search enables filtering for metadata. We also set an example for usability studies of search engines in the digital humanities, using a thinking-aloud set-up and desktop activity recording.

Our recommendations for the future development of search engines in the art history domain are: (1) to involve users in the target groups, both for formulating search questions and for evaluating the application; (2) to study the information-seeking behaviour of diverse target groups (researchers, students, tourists) in more detail and investigate the possibility of tailoring search interfaces to the specific target groups; and (3) to extend the topical labelling of images in art history databases, for example through crowd sourcing.

Acknowledgements

The research for this chapter was funded by CLARIN-NL under grant CLARIN-NL-12-022. We thank our colleagues from Huygens ING, the Netherlands Institute for Art History (RKD), and the Art History department of Radboud University for their collaboration.

References

- 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.
- 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.
- Jacsó, Péter (2004). Thoughts about federated searching. *Information Today*: 21(9).
- 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.
- 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.
- 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>
- 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.
- Shokouhi, M., & Si, L. (2011). Federated search. *Foundations and Trends in Information Retrieval*, 5(1), 1–102.
- Spink, A. (2002). A user-centered approach to evaluating human interaction with web search engines: an exploratory study. *Information processing & management*, 38(3), 401–426.
- Stehouwer, H., Durco, M., Auer, E., & Broeder, D. (2012). Federated search: Towards a common search infrastructure. In *LREC 2012: 8th International Conference on Language Resources and Evaluation* (pp. 3255–3259). European Language Resources Association (ELRA).
- Trant, J. (2009). Tagging, folksonomy and art museums: Early experiments and ongoing research. *Journal of Digital Information*.

- Tunkelang, D. (2009). Faceted search. *Synthesis lectures on information concepts, retrieval, and services*, 1(1), 1–80.
- 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.
- 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.
- 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.