Centrum Wiskunde & Informatica

REPORT*RAPPORT*

INS

Information Systems

INformation Systems

Supporting subject matter annotation using
heterogeneous thesauri, a user study in Web data reuse

M. Hildebrand, J.R. van Ossenbruggen, L. Hardman,
G. Jacobs

Centrum Wiskunde & Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

Supporting subject matter annotation using heterogeneous thesauri, a user study in Web data reuse

ABSTRACT

We performed a user experiment in which museum professionals used vocabularies from the Web for annotating the subject matter of museum objects. We study the requirements on the underlying RDF dataset, search algorithms and user interface design in a real world setting. We identify the advantages of reusing vocabularies from the Web and discuss how and to what extent the disadvantages can be overcome. The study is performed at the Print Room of the Rijksmuseum Amsterdam, where currently a large collection of prints, photographs and drawings is being catalogued. We report on the analysis of the current practice of professional cataloguers, the iterative design of an annotation tool and a qualitative evaluation of this tool with a user experiment in a realistic annotation environment. We discuss our findings in terms of their impact on the RDF data, the semantic search functionality and the user interface.

Supporting subject matter annotation using heterogeneous thesauri
a user study in Web data reuse

Michiel Hildebrand[*,a], Jacco van Ossenbruggen[a,1], Lynda Hardman[a,2], Geertje Jacobs[b]

[a]*Centrum voor Wiskunde en Informatica, Sciencepark 413, 1098 SJ Amsterdam, The Netherlands*
[b]*Rijksmuseum Amsterdam, The Netherlands*

Abstract

We performed a user experiment in which museum professionals used vocabularies from the Web for annotating the subject matter of museum objects. We study the requirements on the underlying RDF dataset, search algorithms and user interface design in a real world setting. We identify the advantages of reusing vocabularies from the Web and discuss how and to what extent the disadvantages can be overcome. The study is performed at the Print Room of the Rijksmuseum Amsterdam, where currently a large collection of prints, photographs and drawings is being catalogued. We report on the analysis of the current practice of professional cataloguers, the iterative design of an annotation tool and a qualitative evaluation of this tool with a user experiment in a realistic annotation environment. We discuss our findings in terms of their impact on the RDF data, the semantic search functionality and the user interface.

Key words: Heterogeneous data, data integration and reuse, semantic annotation, cultural heritage

1. Introduction

We report on a user study that investigates how museum professionals search for appropriate terms within multiple thesauri during an annotation task. The study was performed within the *Print Room Online* project at the Rijksmuseum Amsterdam for a period of 11 weeks, and includes a field study to gather information about the current annotation practices, the iterative design of a prototype interface to support annotation of subject matter and a user experiment to test the final prototype. We discuss the outcome of this study in terms of the requirements on the underlying RDF data, the application's search functionality and user interface design.

Our prototype can be seen as an example of an application that reuses available Web resources and re-purposes rich and highly heterogeneous linked data to support users in a specific task. Although our insights are collected in very specific domain and for a specific task, our observations can be generalized in two ways.

Firstly, to annotation scenarios at other museums, (audio/video) archives and libraries, as many issues also apply to their subject matter annotation tasks. Secondly, to other scenarios in which the reuse of Web data should aid the end user, as the issues we tackle are likely to occur in Semantic Web applications dealing with heterogeneous data. We generalize our findings on the needs for information disambiguation, alignment, multilingualism, compound query support and result visualization and organization to make them relevant for a wider range of applications that reuse Web resources and/or Semantic Web technology.

The paper is organized as follows. In Section 2 we document the current annotation practice at the *Print Room Online* project of the Rijksmuseum. We discuss other approaches to thesaurus-based annotation in Section 3. In Section 4 we sketch the phases of the study that are covered in the following sections: we identify the requirements for subject matter annotation in Section 5, refine these requirements by process of a iterative user interface design in Section 6, and test the resulting prototype in a user experiment discussed in Section 7. We present conclusions and future work in Section 8.

[*]Corresponding author. Phone: +31 (0)20 592 9333; Fax: +31 (0)20 592 419

Email address: `michiel.hildebrand@cwi.nl` (Michiel Hildebrand)

[1]Jacco van Ossenbruggen is also affiliated with VU University Amsterdam.

[2]Lynda Hardman is also affiliated with University of Amsterdam.

2. Current annotation practices at the Rijksmuseum

The Print Room of the Rijksmuseum in Amsterdam, the Netherlands, has a collection of about 700.000 prints, drawings and photographs. Within the project *Print Room Online* the Rijksmuseum aims to register the basic properties of each print, such as the object ID, storage location, title, creator and measurements. In addition, the museum aims to make the collection accessible to the public by making high quality digital scans and adding subject matter annotations. The latter refers to the description of what is depicted on a print and is the focus of this paper. The upcoming three years the project will catalogue 100.000 objects and make them accessible through the museum's website, www.rijksmuseum.nl.

We describe the annotation environment of the *Print Room Online* project and the practices from before the start of our user study. The annotation is performed by seven professional cataloguers. These are highly educated domain specialists, each with knowledge of a particular part of the domain. To improve consistency, the project management has developed an extensive annotation guideline document based on the Spectrum[3] and CIDOC[4] guidelines. The guidelines for the basic registration are straightforward and could be easily applied within the Rijksmuseum's current environment. For the subject matter annotation there is, however, little consensus within the cultural heritage community and very limited tool support. The management team of *Print Room Online* decided to use a temporarily solution. The cataloguers are instructed to describe the depicted person/organization, event name, place and date. In addition, one or more term identifiers from an externally developed art-historic thesaurus (IconClass[5]) should be added when applicable. To save time and to achieve the required throughput rates the subject matter annotation is limited to the main theme being depicted on the object.

Figure 1 shows a schematic view on the annotation fields and thesauri used. The terms used are selected from three internally curated thesauri covering people, places and events. The internal thesauri are stored and accessed in the museum collection management system of which the current annotation facilities are an integral part. Another application, the online IconClass Libertas browser, is needed to search for terms from IconClass.
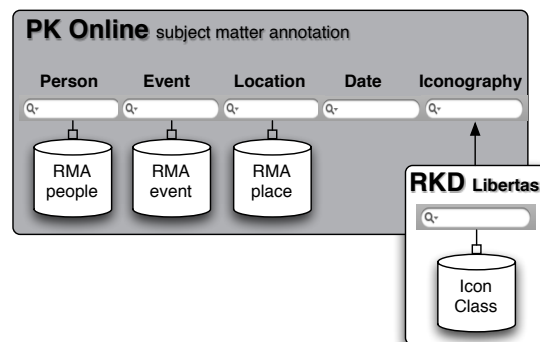
Figure 1: Sketch of the current Rijksmuseum setup for subject matter annotation. Only the five fields that are actually used are shown here. The person, event and location fields each give access to an associated, internal thesaurus. The terms from the external IconClass thesaurus are looked up using another tool, the *Libertas* browser, after which the term identifiers are copy/pasted into the Iconography field. The dates are entered in a free text field that has no associated thesaurus.

	People	Place	Event	IconClass	Total
#	9 245	9 034	6 509	**30 981**	55 796
%	17%	16%	12%	**55%**	100%

Table 1: For each thesaurus, the total number of terms used for the annotation of the Rijksmuseum Print Room objects and the percentage relative to the total number of terms used.

After about one year, 30 747 objects from the Print Room have been annotated. We analyzed the annotations made for these objects. Table 1 shows per thesaurus the number of terms used to annotate the objects. This shows that 55% of the annotation terms used are from the externally developed thesaurus, IconClass. Table 2 shows more details about the usage of the museum in-house thesauri. The table lists the total number of terms in each thesaurus, the total number of terms used for subject matter annotation, the number of unique terms used, the number of terms that had to

	People	Place	Event	Total
terms in thesaurus	65 325	44 541	2 824	112 690
terms used (total)	9 245	9 034	6 509	24 778
terms used (unique)	1 574	1 523	492	3 589
new terms added #	516	169	205	890
new terms added %	**33%**	11%	**42%**	25%

Table 2: Thesaurus terms used for subject matter annotation for the 30 747 Print Room objects. Note the high percentage of missing depicted persons and events that needed to be added.

be added to the thesaurus during the *Print Room Online* project and the percentage of added terms w.r.t. to the unique terms. The table shows that a large number of new thesaurus terms had to be added, in particular for the depicted persons and the depicted event fields. According to the cataloguers, it takes on average about 15 minutes to add a new term to a thesaurus, including the research time. Converted, this means that one person has been adding thesaurus terms for a full month, instead of cataloging objects.

The museum management realizes that curating these thesauri is expensive, and that despite the large efforts, their coverage remains limited. In-house thesauri often reflect the specific perspective of a relatively small group of domain experts, which may cause difficulties when the resulting object descriptions are exposed to a wider user group, for example on the museum's public website. Internal thesauri, including those of the Rijksmuseum, tend to be mono-lingual, which makes the annotations less useful for search applications in a multi-lingual context. Additionally, the quality of the thesauri tends to degrade over time. For example, as different members of the organization add new terms to a thesaurus, the differences in style and the unintended creation of duplicates makes searching for appropriate terms more difficult. Finally, creating a vocabulary with a sufficiently large coverage is virtually impossible when the vocabulary is created by a small group of experts. A result of this is that cataloguers frequently find that the term they need is missing, and are forced to collect the required information before it can be added to the vocabulary. This is a time consuming task that significantly slows down the annotation process.

In contrast to in-house developed thesauri, other, more widely available thesauri[6] are developed and maintained by other parties. These thesauri often reflect the perspective of a broader team of experts, are partially available in multiple languages, are actively maintained with clear quality control guidelines, and, finally, typically provide a much wider coverage of the target domain. The museum often uses these thesauri as a starting when creating new terms for their own thesauri. Recent standardization efforts, such as SKOS [1], significantly lower the technical boundaries to publish thesauri on the Web and reuse them in a specific annotation tool. A drawback of some of these thesauri (e.g.

WordNet) is that they are too general and lack terminology required by a specific museum. A more general drawback of using external thesauri is that the museum loses full control on the content of the thesauri used, and that the thesauri may overlap (that is, the combined thesauri will most likely contain duplicates) and that the thesauri might describe terms from a different perspective than that of the museum.

An annotation tool that would be able to effectively use both internally and externally developed thesauri could, in theory, combine the advantages of both approaches. The key research question is whether we can design an annotation tool in which cataloguers can quickly find an appropriate term from multiple heterogeneously structured thesauri. In particular, we are interested in the requirements on the thesauri and other data needed, on the underlying search algorithms deployed to search multiple thesauri, and on the user interface design, the visualization and organization of the term search results.

3. Related work

Hollink et al. [2] describe an early semantic image annotation tool that supports (subject matter) annotation using terms from different thesauri. An interesting feature is the support for restrictions to limit the search results of annotation fields to terms from specific parts of a thesaurus hierarchy. For example, when searching for terms of a depicted activity only terms from the "activity" branch of WordNet were suggested. Although the tool allowed searching in multiple thesauri for a single annotation field it does not scale well to the number of terms we expect to use for the Rijksmuseum. In particular, we need more scalable visualization and organization of the search results that can also be adapted for the characteristics of specific thesauri.

Finding terms from thesauri is supported in several systems. The Finish Ontology Service Infrastructure, FINNONTO [3], provides several web services on multiple thesauri. One of these services can be used with a client side autocompletion widget to look up thesauri terms. The FACET project also provides an API for different services on thesauri and several interface tools [4]. Both projects provide solutions to access terms from multiple thesauri, but only one thesaurus at a time. Typically the user first selects a particular thesaurus. Our goal is to transparently support search for annotation terms from multiple thesauri.

Generic Semantic Web search engines such as Sindice [5], Swoogle [6] and Falcon [7] give access to

---

[6]Examples include thesauri from the Getty institute, in the Netherlands the artist and art historic (IconClass) thesauri from the Institute for Art History (RKD) and various versions of WordNet in different languages. More details are presented later in the paper.

many vocabularies at once. Often a query leads to several pages of search results, requiring the user to select the most appropriate term. Determining the most appropriate term means, in this case, studying the RDF document a term belongs to, something we can not expect from our end users.

Autocompletion is a technique that continuously provides suggestions while the user is typing. Autocompletion has been applied in applications for many decades. It is in particular successful for "term" search tasks with a limited vocabulary [8], such as email addresses. Autocompletion is also applied to thesaurus term search. Hyvönen et al. provide an overview of different types of semantic autocompletion in [9]. Sinkkilä et al. propose to combine context navigation with autocompletion [10]. They did not experiment with the applicability of semantic autocompletion to end users. In previous user studies we showed that the most suited visualization and organization of autocompletion results differs per thesaurus [11].

For the annotation prototype of the Rijksmuseum our aim is to use autocompletion with multiple large and heterogeneous thesauri to efficiently find terms. This means we need to be able to access multiple thesauri simultaneously, provide scalable search and presentation and configure the visualization and organization of the search results for different types of thesauri.

4. User study

We performed a study at the Rijksmuseum Print Room. The aim of our study is a) to formulate requirements for multi-thesauri term search in subject matter annotation and b) acquire insights into the use of Semantic Web technologies when applied in a real life setting. We decided to collect qualitative information about the annotation practices and test different solutions for multi-thesauri term search. As our study is performed with a small group of experts it is in this stadium not realistic to strive for quantitative data.

The study was performed over a period of 11 weeks and consisted of several phases, as depicted in Fig. 2. The first phase consisted of an analysis of the project's requirements. We present the details of the requirements analysis phase, the main findings and its implications in Sect. 5. Based on these findings we developed a first prototype, which was refined in a process of iterative user interface design. In Sect. 6 we describe the second phase: the implementation of the prototype, the feedback acquired during prototyping, our interpretation of the feedback in terms of design dimensions and the decisions we took. In the third phase we tested our

experimental setup with a pilot study, after which we performed a user experiment in phase 4. We describe the details of the experimental design and the key observations from the experiment in Sect. 7.

5. Requirements analysis

The initial contact with the Rijksmuseum *Print Room Online* project consisted of two sessions in which we acquired information about the current annotation process and formulated the requirements for the prototype. The first session consisted of face-to-face discussions with the project leader, lead cataloguer and a curator of the Rijksmuseum. In the second session, we observed the lead cataloguer annotating several prints during working hours. In addition, we analyzed and discussed the project's extensive cataloguing guideline document and a brief additional document in which the project management sketched their requirements and wishes for subject matter annotation.

5.1. Findings

Due to the extensive guidelines, developed within the *PK Online* project, the basic registration is performed relatively consistent. The annotation of the subject matter, however, remains problematic, as the Rijksmuseum's thesauri do not provide sufficient coverage (see Tab. 2) nor sufficient quality (clogging and limited additional information) needed for adequate annotation. As the project management believes that the integration of externally developed thesauri could improve subject matter annotation in particular, we focused hereon and did not incorporate the basic registration into our prototype.

Based on discussion and observations from several annotation sessions we identified three types of term search tasks relevant for annotation:

1. The user already knows which term to add and needs to (quickly) find this term in one of the thesauri.

2. The user has yet to discover the most suitable term and needs to explore the thesauri to find it.

3. The user suspects a term is not present in any of the thesauri, and needs to confirm this before adding it to one of them.

Our aim is to support all three tasks in a single user interface. Below, we formulate the initial requirements for the vocabulary data, the search functionality and the user interface design.
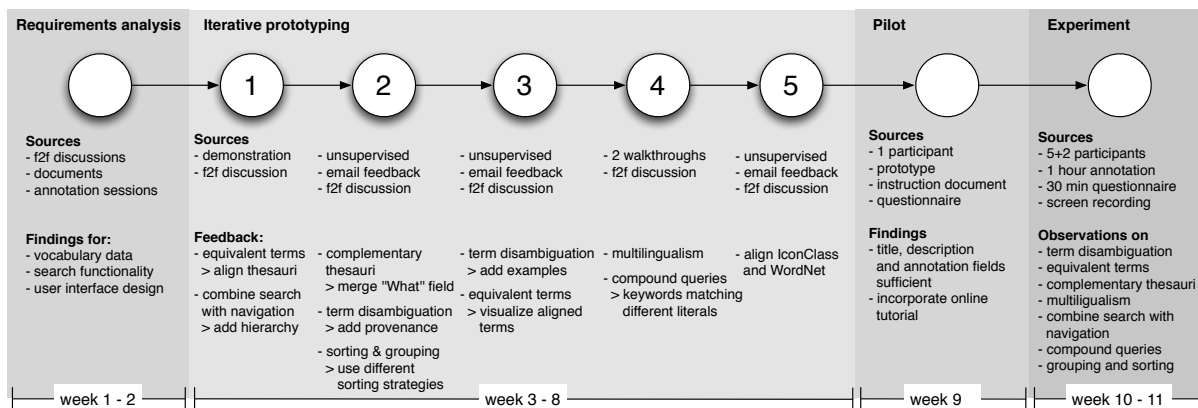
Figure 2: The study at the Rijksmuseum Print Room included a requirements analysis (Sect. 5), an iterative user interface design phase (Sect. 6), a pilot to test the experimental setup and the user experiment (Sect. 7).

5.1.1. Vocabulary data

In the Museum's current collection management system, the annotation fields for person, event and place require terms to be selected from one specific thesaurus. Annotating prints for which all required terms are already in the thesauri is an efficient process, with most of the time being spent on researching what is being depicted, and relatively little time spent on actual data entry. This picture, however, changes dramatically when terms are missing. First, the cataloguer needs to do an exhaustive term search to be sure the term is really missing. Then she needs to research and formulate a request to add the term to the thesaurus, detailing exactly what term needs to be added, along with the additional information that needs to be recorded (e.g. biographical data for persons, geographical data for places), and the provenance data of the literature sources upon which this information is based. When the request has been filed, the cataloguer continues her work, but needs to remember to come back to the annotation record of the associated artwork to add the annotation once the missing term has been added to the thesaurus. The whole process is extremely time consuming and disrupts the normal annotation work flow. Not surprisingly, the project would like to include additional thesauri to increase the term coverage for all three fields currently associated with the internal thesauri (depicted persons, events and places).

One externally developed thesaurus, IconClass, is already used for the iconography annotations. Surprisingly, this thesaurus has not been integrated into the museum's annotation interface or collection management system. Instead, the cryptic IconClass term identifiers (e.g. 45K21) are looked up in an independent web-based interface for IconClass and copy/pasted from the web browser into the annotation field or typed in manually. While cataloguers regularly make mistakes during this process, it is hard to detect such errors because the current tool does not contain the textual labels associated with the cryptic term identifiers.

The terms from IconClass are well suited for the annotation of biblical and mythological stories depicted on museum objects. For annotating more general objects and concepts depicted on other prints and photographs, a more general thesaurus would be required. Since such a thesaurus is currently not available within the project, such annotations are either omitted or added to a free text description field, thereby losing all advantages of thesaurus-based annotation. While for the depicted persons, events and places, the project would like to add external thesauri to increase coverage (e.g. having more terms of similar nature), here an additional thesaurus would need to provide a different type of term (e.g. general terms instead of specific terms).

5.1.2. Search functionality

In the annotation fields of the museum's current system, cataloguers use keyword search to find terms from a specific thesaurus. A single query gives access to a separate page with matching terms from a single thesaurus, which is directly associated with an annotation field. In the current interface, many queries already yield long lists of results, and cataloguers fear this will only get worse when more thesauri are added. The long lists make even the first search task, finding a known term, relatively difficult.

The current tool provides no other means to access a thesaurus beyond keyword search. This limits the

cataloguer with the second term search task, when the right annotation term is not known in advance and the cataloguer has to discover the most suitable term. In a cleanup process of the RMA PEOPLE[7] thesaurus, the Rijksmuseum staff removed many duplicate artist names and added mappings to non-aligned equivalent terms. The occurrence of these duplicate is an indication that the third task, confirming that a term is not present, is also not well supported.

5.1.3. User interface design

The annotation interface of the museum's current system consists of many tabs, showing a maximum of 10, each containing many annotation fields. For the annotation of the depicted subject matter, within the project only five out of the 33 available subject matter annotation fields are actually used. The project management expressed the need to simplify the annotation interface by significantly reducing the number of annotation fields and improving the layout.

The search results in the Rijksmuseum's current annotation interface are visualized by simply showing the term itself. Details about a term are only available in a separate window that is shown on request. Comparing terms within the search results is made more complicated with this type of result visualization, as it requires multiple clicks. What information should be used for visualization depends on the thesaurus. For example, in the IconClass browser, used by the Rijksmuseum cataloguers, the term identifier is also shown in the search results. This identifier indicates where in the hierarchy the term occurs, which helps experienced cataloguers to quickly determine if a result is the right term. For other thesauri, completely different information might be more suited for the visualization of search results.

5.2. Implications

The findings above were discussed and translated to design decisions for the initial prototype in terms of the vocabulary used, the supported search functionality and the interface design.

5.2.1. Vocabulary data

After discussion with the project manager and lead cataloguer, we decided to build a prototype which integrated data from five external thesauri, in addition to the three internal thesauri developed and maintained by the Rijksmuseum. Our aim was to cover different aspects
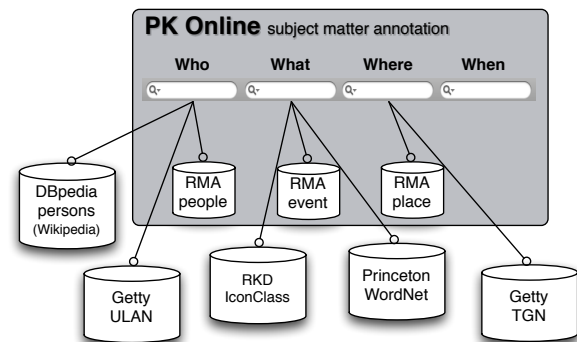

Figure 3: Prototype setup of the subject matter annotation fields and the multiple thesauri used for each field. The initial prototype contained two extra fields (iconography and event) which were later merged into a single *What* field. As in the original annotation interface, the dates are entered into a free field.

with multiple thesauri. As the quality of the data is important for the museum we only integrated internationally respected sources. Figure 3 shows the thesauri used per annotation field. We decided to use two additional thesauri with individuals: Getty's United List of Artist Names[8] (ULAN) and DBPedia's RDF version of person data[9] from Wikipedia; and one additional source of place names: Getty's Thesaurus of Geographic Names[10] (TGN). We also added the RKD IconClass thesaurus and, as a source for more general terms, W3C's RDF version [12] of Princeton's WordNet[11]. An additional thesaurus with relevant historical events and a thesaurus dedicated to persons depicted on portraits was high on the project's wish list but were not available during this study.

5.2.2. Search functionality

As we decided to provide autocompletion suggestions to the user, the search algorithm should support fast prefix search. To support search within multiple thesauri the search algorithm of the prototype should be able to cope with the differences among the thesauri and allow different search strategies to be configured for each thesaurus. The interface will contain different annotation fields that should give access to different types of thesauri terms. For example, only locations should

---

be suggested in the annotation field for depicted locations. To select terms of the right type some form of result filtering is, thus, required. In the RMA, Getty and DBPedia thesauri type information about the terms is available, whereas in WORDNET and ICONCLASS it is not directly indicated if terms are persons, locations or concepts. It is thus not straightforward to filter out the different resource from WORDNET and ICONCLASS.

5.2.3. Initial interface design

The project management suggested that the initial prototype interface should focus on the *Who*, *What*, *Where* and *When* of an object. Only the active fields of museum's annotation interface (5 out of 33): *iconography*, *person*, *event*, *place* and *date* were incorporated into the prototype. An extra field was required to add more general terms about *what* is depicted.

To allow more efficient search in the now much larger set of vocabularies, we decided not to use the current tool's interface where each term search results in a separate "screen" with matching results to choose from. Instead, we designed the prototype interface around annotation fields with "autocompletion". For the purpose of annotation this means that thesaurus terms can be suggested directly within the annotation field, allowing the user to quickly try alternatives and create an annotation with a minimal number of interaction steps. Part of this study is to investigate if autocompletion can support all three term search tasks. Since the project did not use a controlled vocabulary for the date field, nor wished to do so, we did not focus on the interface to enter dates and continued to use a free text annotation field.

6. Refinement of requirements and design decisions

Based on the initial requirements we developed a first prototype, which was refined through an iterative process of redesign and feedback by the project members. All five prototypes were web applications accessible in a standard web browser. This allowed the project members to use the prototype without supervision, in their own time and environment. The first prototype was demonstrated and discussed face-to-face with the project leader. The second and third prototypes were explored unsupervised by the project leader and lead cataloguer, who provided feedback by email and afterwards the findings were discussed face-to-face. The fourth prototype was used for an interactive walk-through by two professional cataloguers, who gave feedback during the walk-through. The final prototype was again explored unsupervised by the project leader and lead cat-

aloguer, with feedback by email and face-to-face discussion. We provide some brief information about the implementation and illustrate the main functionality of the final prototype before explaining the refinements of the requirements and our design decisions.

6.1. Prototype implementation

The client-side interface is implemented in HTML and JavaScript. The interface widgets are developed on top of the Yahoo! User Interface (YUI) library[12] and our configurable autocompletion widget, described in [13]. The server-side search algorithms are implemented in SWI-Prolog using its Web and Semantic Web libraries [14]. The search functionality is accessible through an API over HTTP. A request consists of one or more keywords and a parameter list with search, organization and visualization options. The server returns a structured response in JSON notation that contains matching thesaurus terms and the requested display information. All software is distributed as part of ClioPatria, the open source framework developed as part of the MultimediaN E-Culture project [15].

We required RDF versions of all thesauri to be able to use them in our server middleware. The Getty and the Rijksmuseum thesauri were already converted to RDF within the MultimediaN E-Culture project [16]. Some additional mappings were made to match the SKOS standardization. W3C's version of WORDNET and the DBPedia person data were already available in RDF. For ICONCLASS we used the RDF version developed by the STITCH project[13].

A screen shot of the annotation interface of the final prototype[14] is shown in Fig. 4. The interface has a two column layout. The left side includes an editable title, an image (if available) and a description of the current object. The right side includes the subject matter annotation fields. Each annotation field consists of a header with the name of the field and a brief description of the available terms. Below each header comes a text-input field and a list of the annotations that have been added. The annotation fields for *Who*, *What* and *Where* use autocompletion to suggest terms while the user is typing. An annotation is made by selecting one of the suggestions. An annotation can be removed by clicking the delete button (cross) on the right. All changes directly update the data stored in the back-end. Selecting the

---

[12] developer.yahoo.com/yui
[13] www.nwo.nl/catch/stitch
[14] e-culture.multimedian.nl/pk/annotate?uri= RP-P-OB-77.320

Figure 4: Interface layout of the final prototype. In the left column an editable title field, image and editable description field. In the right column the four subject matter annotation fields. Image of the print used with permission, courtesy of the Rijksmuseum Amsterdam.

link labeled "done" brings the user back to an opening screen, where the annotation of a new object can be started. The link labeled "cancel" does the same and, in addition, removes the annotations already made.

6.2. Feedback

The iterative prototyping and the feedback provided by the project members identified a number of topics that we used to refine the requirements from the first phase. We highlight several issues that we believe are relevant to Semantic Web technologies in general. Figure 2 provides a chronological overview, listing the key feedback topics per prototype. Below we explain for each topic the relevance to Semantic Web technology, the practical problem encountered in our study, our interpretation in terms of the requirements and the design decision we made.

6.2.1. Term disambiguation

An important motivation for the use of unique identifiers (URIs) for resources on the Semantic Web is to solve ambiguity. Each URI only refers to, for example, one person, location or concept. For annotation the user has to choose which URI is the most appropriate for the current task. A label attached to a URI might not be sufficient to determine this — the ambiguity of the labels was why we needed unique identifiers in the first place. WORDNET, for example, contains many homonyms, that is, words that have different meanings. The RMA PEOPLE and ULAN thesauri contain many different resources with the same or a very similar name.

As in many Semantic Web applications, we thus have to decide what information about each resource should be used to present the resource unambiguously. Even when only a single matching term is found, presenting such extra information can also help the user to confirm that this term is indeed the one they had in mind. During the discussion of the first prototype it became clear that for each field, different types of additional information should be shown in addition to the term labels. Figures 5 and 6 show, for example, the different information used for presenting suggestions for the *Who* and

8

Where fields. As already mentioned in the requirements, the cataloguers often use the cryptic identifiers (called *notations* in SKOS) of the IconClass terms, and during the walk-throughs with the fourth prototype it became clear that for disambiguating persons the birth and death dates are very important clues. After the second prototype the cataloguers also requested the addition of provenance information: *"Always indicate where a term comes from. This could be important when deciding on a particular term."* In later prototypes the cataloguers requested even more information to be shown, such as thumbnails with examples of other objects annotated with the term.

Figure 5: Descriptions of individuals in the *Who* field. Underneath the name a short biography is displayed. This contains the nationality, role/profession and birth/death date. Note that for the first person listed, only the profession is available in the data. The abbreviation RMA, shown to the right, indicates the thesaurus source.

Figure 6: Descriptions of locations in the *Where* field. The name of the location is shown, in addition to the associated country. Underneath the name the place type is displayed. Again, the thesaurus source is shown to the right. The first term occurs in both RMA PLACES as well as Getty's TGN, the latter providing most of the additional information.

On the data level, no changes were required to support this request (although the thesauri do not always contain the necessary extra information for each term).

The required extensions to the underlying search functionality were more extensive. For the different fields and different thesauri we needed to include different types of information. Some of the required information was not covered by the standard SKOS model, and some information required some extra computation

on the underlying data. To flexibly support thesaurus-specific configurations, we created a plug-in model on the server executing the generic search algorithm. Plug-ins were used to define the specific information that should be collected for each search result and to compose, for example, a short biography for an individual out of the nationality, role/profession and birth/death date.

In the user interface, the amount of extra information requested by the cataloguers resulted in more information than would naturally fit into the autocompletion suggestions list. We only put what we deemed the most important information in the primary list. The remaining items were shown in a secondary panel, displayed when a term is highlighted. This panel shows, where available, a short description, examples of other objects annotated with that term and the relevant part of the thesaurus hierarchy (see right hand side Fig. 7).

6.2.2. Equivalent terms

In a setting where multiple data resources are aggregated from the web it is very likely that the data contains duplicate resources. In this specific context this means that equivalent terms are found in multiple thesauri. When project members tried out the autocompletion in the first prototype, having the same label occurring multiple times was found to be very confusing. We also observed that the presentation of alternative labels for equivalent terms took up valuable real-estate in the autocompletion result list. We decided that the search results should contain only a single suggestion for each set of equivalent terms. To present the search results from multiple sources in such a way it is thus very important that equivalent terms are aligned.

On the Semantic Web, this means we have to determine if different URIs refer to the same resource. For ontologies and vocabularies this is known as an alignment problem. In our prototype, alignments between the equivalent terms within and across thesauri were required to prevent duplicate results in the interface. Note that alignments are normally used to extend the number of search results by allowing results indexed with terms from one vocabulary to be found with terms from another vocabulary. In contrast, we need the alignments to reduce the number of duplicate search results.

The original RMA thesauri already contained multiple terms for the same person, location or event, with the alignment relations between them. The first prototype did not use these alignments. We corrected this oversight in the next version. We also had to create alignments between the individuals in RMA PEOPLE and Getty's ULAN and the places in RMA PLACES and

Getty's TGN. Our relatively simple mapping tool only aligned identical terms (using `skos:exactMatch` relations), ignoring potential broader or narrower relationships across thesauri.

In the underlying search functionality, we initially extended the search algorithm to use the alignment relations to filter out duplicate terms. On the interface level, however, this yielded unexpected results. Since for each suggested term extra information is displayed (e.g. biographical information, short descriptions, alternative terms), removal of duplicates also resulted in the loss of such extra associated information. The project members had several requests and questions concerning this when confronted with the next prototype: *"What happened to the terms left out? We can't see all the names. [...] Some terms from the RMA thesauri have limited additional information, is it possible to enrich this with information from equivalent terms from other thesauri?"*

To address this, we again modified the underlying search algorithm, but now to use the equivalence relations to collect, for each result, all relevant information on its equivalent terms. This solution allowed the user interface to present a set of equivalent terms as a single result, but also to use the extra information available from all the thesauri. We also wanted the interface to show only one preferred label for such a set of equivalent terms, but different thesauri often indicate different preferred labels for the same concept. We solved this ad-hoc for this specific project, because the members expressed a clear preference for the labels in the Rijksmuseum thesauri.

6.2.3. Complementary thesauri

Some data sources available on the Web can be naturally combined for a specific function. For example, in the prototype the different thesauri with individuals were all used in a single annotation field to describe depicted persons. For other sources such a combination may seem less obvious. In the initial prototype WordNet, IconClass and the RMA Event thesaurus were all accessible from separate annotation fields. After the cataloguers tested the second prototype it became clear that the difference between the WordNet and IconClass fields was unclear. *"The distinction between these two fields is not intuitive for the cataloguer. One term can occur in both the* What *and the* Iconography *fields. Perhaps it is easier for the cataloguer if the information is presented in one field, but with a distinction between the* IconClass *and* WordNet *terms."*. Different thesauri may be complementary to each other for a specific task. In our prototype the advantage of combining comple-

mentary thesauri in a single annotation field is that alternative suggestions from different thesauri are given simultaneously for a single query.

The cataloguers indicated that the *What* field should provide access to IconClass, WordNet and the RMA Event thesaurus simultaneously. To realize this request, on the data level we added a superclass containing the terms of all three thesauri, since the search algorithm already supported subclass reasoning in the filtering of the search results. In the user interface, we configured a single autocompletion field to search all three thesauri.

6.2.4. Multilingualism

The Web contains sources in different languages. Ontologies and vocabularies may contain labels and descriptions in multiple languages. Limiting the data sources used in an application to a particular language may, however, rule out many useful resources. The RMA thesauri used in the prototype were all in Dutch, but we only managed to find additional sources in English. Even IconClass, that was originally developed in the Netherlands, is not yet available in Dutch. Feedback on the third prototype indicated that: *"The different languages in the data could cause a problem. For example when searching in the* What *field it is not possible to search for* siege *and find an equivalent Dutch term from the* RMA Event *thesaurus."*

We decided not to change the data or the underlying search functionality, but to use the experiment to explore the practical implications of having both Dutch and English thesauri in a single annotation field. Given the expertise of our users, we only changed the user interface to briefly indicate in the description of the annotation field headers which language to use for which thesaurus (see the header of the What field in the top right of Figure 7 for an example).

6.2.5. Combining search with navigation

When searching for resources on the Semantic Web the user may not always know in advance what exactly she wants to find. Instead she might prefer to explore the available data to determine which resource is most suited for her. For explorative tasks, keyword search can be a good starting point, but not always sufficient. When trying out the first prototype the cataloguers were pleased with the autocompletion functionality, as it allowed them to quickly find known terms from multiple thesauri. When using it to search in IconClass, however, some functionality compared to the online IconClass browser was missing. The cataloguers often perform a global search to find a relatively generic term, which they use as a starting point for further navigation

along the hierarchical structure to find a more specific term. This helps them to find suitable terms without knowing the exact term in advance.

To support a similar type of functionality we decided to, on the data level, use the `skos:broader` relation to model all hierarchical relations among terms. The RDF version of IconClass already used this relation to model its concept hierarchy. For WordNet, we mapped its lexical hypernym relation, and for TGN and RMA Places we mapped their geographical containment relation to `skos:broader`.

In the underlying search functionality, we created a separate API that provides, on request, the data for the partial hierarchy as well as data about the children when the hierarchy is further expanded. Using this, we extended the user interface to show parts of the hierarchy for each autocompletion suggestion. The screen shot in Fig. 7 shows the hierarchy for the highlighted suggestion in a secondary panel. The hierarchy contains the term itself, all its ancestors and the direct children. More descendants are available on request by further expanding the direct children.

6.2.6. Compound queries

When searching on the Semantic Web with a compound query, the query could match a single literal containing all these keywords, or match multiple literals of different related resources. In our study we were confronted with both these variants of compound queries.

First, during the walk-through of the fourth prototype we noticed that when searching for specific thesauri terms, such as the historical and religious events in Icon-Class, a single keyword is often not enough. For example, there are 490 terms in IconClass matching "Mary". Adding an additional keyword can greatly reduce the number of search results, for example, when the query "Mary" is extended with the keyword "assumption" only 5 results are left. The IconClass browser and the current annotation interface of the Rijksmuseum did not support queries consisting of multiple keywords. the cataloguers would, therefore, first search on a generic term and then navigate to the more specific term. Often the cataloguers, however, knew the exact term or at least multiple keywords contained in the term. In the search algorithm we added support to match multiple keywords within a single literal.

Second, In the *Where* field of the fifth prototype we extended the algorithm to also match multiple keywords against different locations and use the hierarchical relations to find the best term. This allows the user, for example, to add the name of the country to a query for a

city name. At the data and interface level no additional support was required for compound queries.
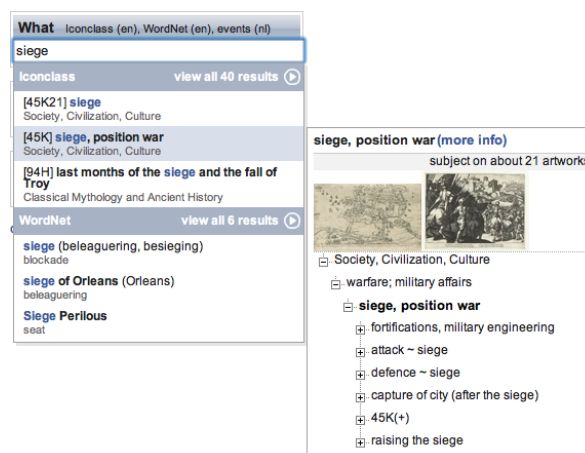
6.2.7. Sorting and grouping search results

Figure 7: Annotation field with autocompletion. Suggestions are shown for the query "siege". Results from both IconClass and Word-Net are shown (left), each presented in a separate group. A secondary panel (right) shows more information for the highlighted term ("[45K] siege, position war"), including the hierarchical structure the term is part of. The hierarchy contains the term itself in bold, its ancestors and its direct children.

When a search application provides many search results for a query, some form of organization of these results is required to support the user with finding the right result. Ranking search results is an efficient technique, but can only be applied if data provides a good criterion to rank on. Initially, we sorted the search results on the frequency of use, showing those used most frequently for annotation as the first autocompletion suggestions. After trying the second prototype the cataloguers, however, indicated that: *"alphabetical sorting would be better for individuals and events"*. During the discussion they indicated that alphabetical sorting helps them to quickly scan a list of names.

Alphabetical sorting is useful where there is a relatively large number of results and the cataloguers have to find the right term among them, or to determine that a term does not exist. In these tasks a frequency ranking is not helpful as it does not give any insight about where a term can be found in the list. A potential problem with alphabetical sorting is that the format of the labels can vary among terms. For example, in the RMA People thesaurus the preferred label consists of the forename and then the surname, while it is reversed in ULAN. We decided not to change the original names in the thesauri and accept the effects on the sorting strategy.

In the *Who* field it is common to have a large number of search results, as there are many historical individuals with a similar name. At the same time, there are still many individuals not contained in any of the thesauri. To support the cataloguer in finding terms from a long list, or determining that a term is missing, we used alphabetical sorting in the *Who* field. In the *Where* field the opposite was the case. The two place thesauri together provide high coverage for the annotation task at the Rijksmuseum, and there are few duplicate place names[15]. We thus ranked the places according to their frequency of use.

In the *What* field we needed three different sorting strategies. The cataloguers were used to navigating the IconClass hierarchy starting from the highest matching term. We thus decided for IconClass to show the terms highest in the hierarchy first. For WordNet, we used the frequency counts stored in this thesaurus, which indicate the relative frequency of use among homonyms in English. For the events in the Rijksmuseum thesaurus we used alphabetical sorting.

To support the different sorting strategies we used the same plug-in mechanism as for the result visualization. The plug-ins define which information should be used for sorting, which is then used by the algorithm to sort the terms. Because all the sorting is done by the underlying search algorithm, no further changes were needed in the user interface implementation.

Grouping similar types of results is another organization strategy that is often applied to search results. Grouping has the advantage of displaying a wider variety of choices in the same screen real estate. In the prototype the cataloguers wanted to search simultaneously on multiple complementary thesauri in the *What* field and at the same time keep a clear distinction between the terms from each thesaurus. We decided to create the distinction between the thesauri terms by visually grouping the search results coming from the same thesaurus together.

On the data level, we needed to add the `skos:inScheme` property to each term to explicitly specify to which thesaurus it belongs. In the underlying search functionality, we extended the search algorithm and API with the option to group results by any property. In the user interface the different groups were realized by adding group headers to the result list, as shown in Figure 7. For each group a maximum of three suggestions were shown. Clicking the group header

allowed the user to view all suggestions within that group.

7. Evaluation

The goal of the user experiment was to qualitatively evaluate the solutions proposed in the prototyping phase, by asking all cataloguers to use our prototype to annotate a number of "new" artworks (that is, artworks they have not annotated before) from their own area of expertise. To test our design for the user experiment we first performed a pilot with one cataloguer from the Rijksmuseum.

Our aim was to use the final prototype in a realistic environment. In our initial design the cataloguer would start describing the object using their own annotation interface and switch to the prototype for the subject matter annotation. During the pilot it became clear that only the editing of the title and description interacted with the subject matter annotation. The cataloguer would, for example, start with a description, make some annotations and then realize the description should be changed. The cataloguer also used the autocompletion suggestions to find the correct spelling for names she wanted to use in the title. Halfway during the pilot study the project leader of *Print Room Online* suggested that, for a realistic environment, it would suffice to enter also the title and description fields in the prototype, and not to use the museum's own annotation interface during the experiment at all. The second half of the pilot was successfully continued using the simplified setup.

In the pilot it also became clear that some functionality that was supported by the prototype was not noticed by the cataloguer. We decided to add an online tutorial, at the start of the experiment, to acquaint the participants with the supported functionality. Below, we first describe the design of the experiment after which we present the general observations on the coverage of the date used and the use of autocompletion. Second, we give a qualitative evaluation of the design decisions made in the prototyping phase.

7.1. Experimental design

The participants of the experiment consisted of five professional cataloguers and two museum professionals who occasionally create annotations. Each participant took part in a one hour annotation session, annotating about six "new" objects. The objects were chosen by the project leader and matched the cataloguers expertise. Before the session the participant read an

---

[15]Note that the North America section of TGN contains many places with the same name. Because items from this part of the world are vary rare in the Print Room, we decided to omit this part of TGN.

instruction manual[16] and went through the interactive tutorial of about 15 minutes. After the annotation session they filled in a questionnaire[17]. We asked a similar set of questions for each annotation field. The questionnaire focused on the topics for which we had provided solutions in the prototype phase. For example, to test term disambiguation we asked the participants: "How confident were you that the selected term was the term you intended?", "Was it clear for each suggestion from which thesaurus it came?" and we asked them to rank the different types of additional information that were presented. Further questions were about the usefulness/annoyance of the (missing) alignments between equivalent terms, usefulness of autocompletion, the sorting and grouping strategies and the formulation of compound queries. In addition we asked the participants for demographic information and their experience with autocompletion.

Besides the questionnaires we used two other sources for evaluation. First, the observation of the annotation sessions. All sessions were screen captured and observed in real time by two researchers. The captured videos were annotated, focusing on the query construction and the result selection. Second, query logs showed precisely which characters were typed in the autocompletion fields and which annotations were made. The results that we present are based on the combined analysis of the questionnaires, observation notes, screen recordings and query logs.

7.2. Qualitative evaluation of design decisions

The key findings of the evaluation are summarized in Tab. 3. Below, we discuss the findings for each design problem.

7.2.1. Term disambiguation

In the questionnaire we asked the participants to rate the usefulness of the additional information for the different types of terms. As we only recorded the opinion of seven participants we do not make any statistical claims about the importance of particular types of information. We merely want to illustrate that cataloguers prefer different types of auxiliary information for different types of terms. For individuals the short biography is always considered the most important type of information. In particular, we observed that the cataloguers used the birth and death dates to compare against the

Design problem	Findings
Term disambiguation	Different types of terms require different types of additional information to disambiguate them.
Equivalent terms	Conservative alignment is important for annotation to remove duplicates while preventing false positives.
Complementary thesauri	Simultaneous search in complementary thesauri helps cataloguers to choose the most suited term as they can directly compare alternatives.
Multilingualism	Professional cataloguers can deal with sources in multiple languages. Autocompletion helps to deal with multilingualism, as queries in different languages can be tried quickly.
Combine search with navigation	Autocompletion helps in annotation, as users often have to extend a query and try multiple different queries. Autocompletion can be successfully combined with a partial hierarchy to support selection of more specific terms. For some cases it might be useful to have access to a full search/navigation interface.
Compound queries	To find specific thesaurus terms from a set of very similar terms the user should be able to specify a query with multiple keywords.
Sorting and grouping	Professional cataloguers prefer a transparent sorting method, such as alphabetical sorting. Depending on the type of terms a frequency ranking might help for less experienced cataloguers. Within a single annotation field it helps to distinguish terms with a clearly different type.

Table 3: Summary of the key findings of the user study with respect to the seven design problems discussed in Section 6 and 7.

creation date of the print. The description and alternative spellings are also considered useful. For locations the place type, the hierarchy and the description are all considered important. Hierarchy information is also rated as very important for IconClass, but not at all for WordNet. For the latter, the description and synonyms are preferred.

During the prototyping the project leader and lead

---

[16] e-culture.multimedian.nl/rma/prototype_manual.pdf
[17] e-culture.multimedian.nl/rma/questionnaire.pdf

cataloguer both stressed the importance of the provenance information. Most participants, however, stated in the questionnaire that they were not interested in seeing the provenance information. *"I don't care much about knowing where a term comes from. I just want the right term (=most specific)."* [P4][18]. Our explanation for this discrepancy is that during the prototyping the provenance was crucial to get an idea about the added value of the additional different thesauri. The lead cataloguer was also very much interested in seeing which terms were missing from their own thesauri to assess their quality. For the actual annotation tasks, however, this turned out to be less important than we had expected. Furthermore, the cataloguers did not consider if the thesauri were from authoritative sources, as they counted on the project management to handle this.

7.2.2. Equivalent terms

Only three cataloguers reported to have seen duplicate terms from different thesauri. While two of them indicated they were not so disturbed by this, the third indicated: *"I would be forced to check both suggestions to see which one is most suited for me – more work."* [P4]. Most cataloguers also didn't notice that some suggestions were, in fact, a combination of multiple aligned equivalent terms and they also indicated not to care about this. We got the impression that the cataloguers are not bothered with the occurrence of incidental duplicates, *"understandable when multiple thesauri are used"* [P6], but that systematic occurrence of duplicates would disturb their efficiency. Aligning equivalent terms is thus important, but does not need to be perfect: a few false negatives result in relatively harmless and occasional duplicates[19].

7.2.3. Complementary thesauri

All cataloguers were pleased with the possibility to simultaneously search in the different thesauri of the *What* field. *"Finding alternatives in* WordNet *is a plus if* IconClass *falls short."* [P1]. WordNet was, in particular, useful to describe *"concrete things"* [P2]. Due to the different perspectives between IconClass, art-historic, and WordNet, linguistic, we did not provide alignments between these two thesauri. In case similar suggestions were provided from both, the participants were instructed to decide per annotation which

thesaurus was most appropriate. In practice, they tended to use IconClass as the primary source: *"I am tempted to select terms from IconClass and use WordNet as a backup simply because they are presented in this order."* [P4].

7.2.4. Multilingualism

As expected, the language variation among thesauri terms required the participants to do some extra work: *"Sometimes a person name is translated from French to Dutch, Louis=Lodewijk. You have to know this."* [P3]. In practice, the cataloguers sometimes had to try queries in multiple languages. In general, the translation itself caused few problems for the professionals. Only for the look-up of English terms from WordNet and IconClass they occasionally had to use a Dutch to English dictionary, which slowed down the process considerably. This is, however, also a problem in the current situation where project members have to search using English terms in the IconClass web interface.

7.2.5. Combining search with navigation

All participants indicated in the questionnaire that autocompletion was "very useful". *"Autocompletion increases speed of working and consistent use of terms."* [P6]. During the experiment we observed that participants often used multiple queries to find a term or to determine that a term does not occur in one of the thesauri. *"You have to figure out yourself how it could already be stored in the thesaurus."* [P5]. The feedback provided by the autocompletion suggestions helped the cataloguers to see that a) there are too many results to investigate, b) the results do not contain the intended term and c) there are no results at all. As autocompletion provides instant feedback for every character typed, the user can quickly switch between scanning the list of results, reducing or extending the query or creating a new query. An example of (a) is that P3 queried on "hendrik" and the system indicated that there were 778 matching individuals. This made her decide to continue typing to specialize the query to "hendrik IV", which returned only 5 individuals. To find a Dutch historical individual, P5 quickly tried different spelling variations of the name *dijck*, *dyck* and *dijk*, using the suggestions to quickly determine that the right term was not found (b) or that there were no results at all (c).

The hierarchical structure that was shown for the autocompletion suggestions was used several times to select a more specific term. When P5 investigated the suggestion "peace negotiations" from IconClass the hierarchy contained "signing of peace treaty, concluding the

---

[18]Quotes from the participants have been anonymized and are indicated with P1 to P7).

[19]False positive alignments have a more severe impact. If term A ad B are falsely aligned, they will be presented as a single result, and the user will no longer be able to select one of them. In this scenario, therefore, a conservative alignment approach is best suited.

peace" and this more specific term was selected. Presenting only the partial hierarchy was sufficient to support this interaction. The participants, however, also indicated they wanted to navigate the full hierarchy as provided in the IconClass web interface: *"I missed the overview in the full list with suggestions."* [P1]. *"I would like to see the whole hierarchy, but that could also be because I am used to it."* [P3].

7.2.6. Compound queries

In the questionnaire all cataloguers indicated that the support for compound queries was very useful. A query consisting of multiple keywords allowed them to quickly find specific known terms. This was in particular useful to find specific terms from IconClass, which contains many similar terms about the same topic. *"The Online IconClass browser does not support compound queries requiring more time to find the right term. Now you can find a specific term, such as a biblical event in one go. Very nice!"* [P3].

7.2.7. Sorting and grouping search results

The five professional cataloguers that participated in the study carefully investigated the search results before selecting a term, whereas the two less experienced participants tended to select the first appropriate term. In cases with more than one search result, the professionals scanned the entire list for possibly better candidates. The ranking of the search results is, therefore, for them not so important. *"Ranking is not so important as I can scroll through the list."* [P2]. In fact, the cataloguers prefer the search results to be sorted in a way that is understandable. *"Alphabetical sorting is what I expect for a name list."* [P5]. The alphabetical sorting of the search results that we used for the *Who* field helped the user, as it makes it clear where to look for a term and when to stop looking. Alphabetical sorting also helps the user when not all results are directly shown, as it is clear which results can be expected if more results are requested. The non-alphabetical sorting methods we used for the other annotation fields were judged differently by the participants. They were not particularly liked or disliked.

The participants indicated that some type of grouping of the search results was perceived as useful. In particular, the separate group for events was considered useful: *"Clearly separates the events, very useful when searching."* [P2]. The distinction between terms from IconClass and WordNet was not considered intuitive by all participants. There is some overlap between IconClass and WordNet, and presenting these similar terms

	Who	What	Where	Total
terms used	14	65	16	94
terms not found #	9	9	1	19
terms not found %	41%	14%	6%	20%

Table 4: Per annotation field the number of terms found and not found during the user experiment.

	Who	What	Where	Total
only RMA	13	1	3	17
only external	0	64	4	68
RMA & external	1	0	9	10
total	14	65	16	94

Table 5: Breakdown per thesaurus of the terms found in the experiment.

in different groups was often found confusing. As indicated in the previous observation, seeing the source of the term is not so important to the cataloguer. The positive feedback on the separate event group could be an indication that a grouping by different types of terms is more suitable. This would require, first, to filter out the persons and locations in IconClass and WordNet use these in the *Who* and *Where* fields. Second, the remaining concepts in IconClass and WordNet should then be further classified in a fashion that is logical for the cataloguers, for example, by distinguishing named events from generic objects.

7.3. Term coverage

Table 4 shows the number of terms found and not found in the user experiment per annotation field. These numbers give some insight into the type of terms the cataloguers used during the experiment. Most annotations were added in the *What* field. The weak coverage of the thesauri for depicted people is still present. Table 5 shows a breakdown of the terms found into those found only in an RMA thesaurus, in one of the external thesauri or in both RMA and an external thesaurus. The two external thesauri with individuals (ULAN and DBPedia) did not provide terms useful in this annotation task. For the geographical places the external thesaurus TGN provided additional coverage. Many depicted place names are in both RMA Places and TGN. This overlap can also be considered an improvement, as TGN provides additional information that is lacking in RMA Places. Participants found this additional information helpful in term disambiguation.

15

For the terms needed in the *what* field, the combination of WORDNET and ICONCLASS provided adequate coverage. The lack of an additional event thesaurus is illustrated by the fact that only one term was selected from the RMA EVENT thesaurus. During the experiment we also observed three cases in which an historical event was missing. Note that a relatively high number of terms was selected from WORDNET, an a-typical source in this art-historical context. These were mainly general terms that were either not present in ICONCLASS, or only in a very specific biblical or mythological context. The addition of WORDNET was considered a big added value by most project members.

8. Conclusions and future work

We have derived requirements and design decisions to support a term search task on heterogeneous data in a real life setting. The study sets a first step to picture the landscape of multi-thesauri term search on the Web of Data, which enables further quantitive studies to be performed on thesaurus based annotation.

From the perspective of the museum staff, the experiment was successful. They appreciated the integration of the collection data, internal thesauri and external thesauri into a single tool, the autocompletion search functionality and the extra information that is displayed for each matching term. They also realize, however, that deploying a similar "open" annotation tool in their daily work-flow will require several actions. First, all other tools in the tool chain should be adapted to allow the use of references to external thesaurus terms. This would require support for URIs instead of keywords or internal database keys. Second, it requires a dialogue with other cultural heritage institutions about the organizational aspects, such as the control and maintenance of the data. The museum will have to deal with complex questions about content, organization an technical solutions. This project will be quite a challenge for an art institution. The Rijksmuseum is exploring the possibilities.

The extra thesauri deployed provided mainly a *quantitative* addition, by providing more terms of a particular type. In some cases, the addition was more of a *qualitative* nature, for example where the more general WORDNET terms were better suited to describe photographs than the specific terms from ICONCLASS. The integration of externally developed thesauri also proved beneficial because they provided more information about each term and were multilingual. In the prototype the coverage of persons and events was still limited, as suitable thesauri were not available. A source for depicted persons could be an internal thesaurus from the iconographic institute of the RKD containing Dutch historic figures depicted on portraits.

Some of the disadvantages of re-using data from heterogeneous thesauri can be overcome with careful data alignment and enrichment, suitable search functionality and domain specific configuration of the user interface.

On the data level, thesauri may partially overlap, and we use common vocabulary alignment methods to detect equivalent terms so we could avoid duplicates in the user interface. In our case a conservative alignment method was most suited as wrong alignments are harmful, because they remove possible candidates from the search results whereas a few duplicates in the interface are acceptable. To distinguish ambiguous terms from one another we present each term with additional associated information. For this, mappings to SKOS significantly help to address the heterogeneous structure of the various thesauri, as such mappings yield common properties for preferred and alternative labels, scope notes and broader/narrower relations across all thesauri. In addition, we need extensions to SKOS for representing common biographical and geographical properties. In future work we would like to partially align WORDNET and ICONCLASS and enrich these vocabularies to identify terms describing persons, location names and events.

We require search functionality that goes beyond the functionality of a standard SPARQL endpoint. Fast prefix string matching on RDF literals is required to support autocompletion, along with filtering of the matching results based on the type, source or other properties of the associated resource. Queries with multiple keywords such as "Paris France" require additional support since they potentially match on resources with a label matching one keyword and a related resource matching another keyword. This needs to be configurable by the client as it may be different depending on the task and thesauri used. The search engine also needs to have configurable support to combine several search results related by `skos:exactMatch` relations into a single, coherent search result. Finally, to support hierarchical navigation in the autocompletion we also need efficient ways to retrieve the full path to the root of the hierarchy for each search result.

In future work we would like to improve the search algorithm by suggesting relevant results based on contextual information. One source of contextual information is other thesaurus terms already added to a museum object. Another source could be the free text in the title and the description, where automatic named entity extraction could provide context or be suggested as annotation terms. The contextual information could also

be used to suggest annotation terms.

At the interface, different thesauri require different sorting, ranking and grouping strategies. For the visualization, different types of information and different forms of presentation are required in order to best support the end user in disambiguating and selecting the most appropriate term. In addition to keyword search, some thesauri also require navigation interfaces to explore broader and related terms. In current work we are developing a method that allows the configuration of the visualization and organization by mapping domain specific semantics to an intermediate model.

Acknowledgments

References

[1] W3C, SKOS Core Guide, W3C Working Draft, W3C Candidate Recommendations are available at http://www.w3.org/TR, edited by Alistair Miles and Dan Brickley (November 2, 2005).
http://www.w3.org/TR/swbp-skos-core-guide/

[2] L. Hollink, G. Schreiber, J. Wielemaker, B. Wielinga, Semantic annotation of image collections, in: Second International Conference for Knowledge Capture (KCAP2003) Workshop on Knowledge Markup and Semantic Annotation, 2003.
http://www.cs.vu.nl/~guus/papers/Hollink03b.pdf

[3] E. Hyvönen, K. Viljanen, J. Tuominen, K. Seppälä, Building a national semantic web ontology and ontology service infrastructure—the finnonto approach, in: Proceedings of the European Semantic Web Conference ESWC 2008, Springer, 2008.
http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf

[4] C. Binding, D. Tudhope, KOS at your service: Programmatic access to knowledge organisation systems., J. Digit. Inf. 4 (4).
http://dblp.uni-trier.de/db/journals/jodi/jodi4.html#BindingT04

[5] G. Tummarello, E. Oren, R. Delbru, Sindice.com: Weaving the open linked data, in: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea, Berlin, Heidelberg, 2007, pp. 547–560.
http://iswc2007.semanticweb.org/papers/547.pdf

[6] L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, P. Kolari, Finding and Ranking Knowledge on the Semantic Web, in: Y. Gil, E. Motta, V. R. Benjamins, M. A. Musen (Eds.), 4th International Semantic Web Conference, ISWC 2005, Lecture Notes in Computer Science, Springer, Galway, Ireland, 2005, pp. 156–170.
http://ebiquity.umbc.edu/_file_directory_/resources/118.pdf

[7] G. Cheng, W. Ge, H. Wu, Y. Qu, Searching semantic web objects based on class hierarchies, in: WWW 2008 Workshop on Linked Data on the Web, 2008.
http://events.linkeddata.org/ldow2008/papers/12-cheng-ge-searching-semantic-web-objects.pdf

[8] M. Hildebrand, J. van Ossenbruggen, A. Amin, L. Aroyo, J. Wielemaker, L. Hardman, The design space of a configurable autocompletion component, Tech. Rep. INS-E0708, CWI (November 2007).
http://www.cwi.nl/ftp/CWIreports/INS/INS-E0708.pdf

[9] E. Hyvönen, E. Mäkelä, Semantic Autocompletion, in: Proceedings of the first Asia Semantic Web Conference (ASWC 2006), Beijing, 2006, pp. 739–751.
http://www.seco.tkk.fi/publications/2006/hyvonen-makela-semantic-autocompletion-2006.pdf

[10] R. Sinkkilä, E. Mäkelä, E. Hyvönen, T. Kauppinen, Combining context navigation with semantic autocompletion to solve problems in concept selection., in: K. Belhajjame, M. d'Aquin, P. Haase, P. Missier (Eds.), SeMMA, Vol. 346 of CEUR Workshop Proceedings, CEUR-WS.org, 2008, pp. 61–68.
http://dblp.uni-trier.de/db/conf/esws/semma2008.html#SinkkilaMHK08

[11] A. Amin, M. Hildebrand, J. van Ossenbruggen, V. Evers, L. Hardman, Organizing suggestions in autocompletion interfaces, in: 31st European Conference on Information Retrieval, Toulouse, France, 2009, to be published, based on techreport: http://ftp.cwi.nl/CWIreports/INS/INS-E0901.pdf.

[12] M. van Assem, A. Gangemi, G. Schreiber, Conversion of WordNet to a standard RDF/OWL representation, in: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, 2006.
http://www.cs.vu.nl/~mark/papers/Assem06a.pdf

[13] M. Hildebrand, J. R. van Ossenbruggen, A. K. Amin, L. Aroyo, J. Wielemaker, L. Hardman, The Design Space Of A Configurable Autocompletion Component, cwi INS-E0708, CWI (2007).
http://www.cwi.nl/ftp/CWIreports/INS/INS-E0708.pdf

[14] J. Wielemaker, M. Hildebrand, J. van Ossenbruggen, Prolog as the Fundament for Applications on the Semantic Web, in: Proceedings of the ICLP'07 Workshop on Applications of Logic Programming to the Web, Semantic Web and Semantic Web Services (ALPSWS2007), Porto, Portugal, 2007.
http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-287/paper_1.pdf

[15] J. Wielemaker, M. Hildebrand, J. van Ossenbruggen, G. Schreiber, Thesaurus-based search in large heterogeneous collections, in: A. P. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. W. Finin, K. Thirunarayan (Eds.), International Semantic Web Conference, Vol. 5318 of Lecture Notes in

Computer Science, Springer, Berlin Heidelberg, 2008, pp. 695–708.
http://dx.doi.org/10.1007/978-3-540-88564-1_44

[16] G. Schreiber, A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Ossenbruggen, A. Tordai, J. Wielemaker, B. J. Wielinga, Semantic annotation and search of cultural-heritage collections: The multimedian e-culture demonstrator, J. Web Sem. 6 (4) (2008) 243–249.
http://dx.doi.org/10.1016/j.websem.2008.08.001