

Designing a Thesaurus-Based Comparison Search Interface for Linked Cultural Heritage Sources

Alia Amin¹, Michiel Hildebrand¹, Jacco van Ossenbruggen^{1,3}, Lynda Hardman^{1,2}

¹Centrum Wiskunde & Informatica (CWI), Amsterdam,

²University of Amsterdam (UvA),

³VU University Amsterdam,

firstname.lastname@cwi.nl

ABSTRACT

Comparison search is an information seeking task where a user examines individual items or sets of items for similarities and differences. While this is a known information need among experts and knowledge workers, appropriate tools are not available. In this paper, we discuss comparison search in the cultural heritage domain, a domain characterized by large, rich and heterogeneous data sets, where different organizations deploy different schemata and terminologies to describe their artifacts. This diversity makes meaningful comparison difficult. We developed a thesaurus-based comparison search application called LISA, a tool that allows a user to search, select and compare sets of artifacts. Different visualizations allow users to use different comparison strategies to cope with the underlying heterogeneous data and the complexity of the search tasks. We conducted two user studies. A preliminary study identified the problems experts face while performing comparison search tasks. A second user study examined the effectiveness of LISA in helping to solve comparison search tasks. The main contribution of this paper is to establish design guidelines for the data and interface of a comparison search application. Moreover, we offer insights into when thesauri and metadata are appropriate for use in such applications.

Author Keywords

comparison search, thesauri, cultural heritage

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation

General Terms

Human Factors, Design

INTRODUCTION

In an in-depth study on information seeking needs in the cultural heritage domain, comparison search was identified

as an example of an information seeking task that experts perform frequently for their work [3]. Comparison search involves examining objects or sets of objects for similarities and differences. The more objects there are to compare, and more properties to compare on, the more complex the task becomes. For such tasks, support tools are indispensable. Comparison search tasks are commonly found in the e-commerce domain. For example, a customer who wants to buy a product online might be interested in comparing products from different manufacturers using various properties, such as quality, price, features, and delivery time.

Comparison search also occurs in the cultural heritage domain. For example, consider an art historian doing comparative studies on Dutch paintings owned by different museum collections throughout the Netherlands. First, the historian needs to thoroughly *search* for all paintings made by Dutch artists in different archives and digital museum collections. Second, after all artworks are *selected*, the historian might need to *compare* the artworks by their distinguishing properties, such as by artist, materials used, art style or year of creation, to be able to identify trends. In practice, the scenario above is not trivial. First, there are difficulties in thoroughly searching for artworks in collections. Many museums have their own thesauri from which they use terms to annotate their artworks. As a result, a user needs to be familiar with the different terms in these thesauri and how they are used to annotate the artworks. This is unlikely, as typically, only the museum employees have such detailed knowledge. Second, most museum collections and archives offer poor support for complex search tasks such as comparison search. Most tools used only offer simple interaction, such as keyword search. This is found to be too restrictive for complex tasks such as comparison search [3].

In this paper, we discuss a user-centric approach to support cultural heritage experts in searching and comparing artworks. This research consists of several phases. We carried out a preliminary study to better understand how experts conduct comparison search in practice for their daily work. Based on this study, we derived design requirements and identified key features for a thesaurus-based comparison search tool that supports users to search, select and compare artworks. We then implemented LISA to help experts compare sets of artworks. Finally, we evaluate how well LISA supports the comparison search task. The contribution of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUI'10, February 7 - 10, 2010, Hong Kong, China.

Copyright 2010 ACM 978-1-60558-515-4/10/02...\$10.00

this paper is to establish design requirements for the data and interface of a comparison application. Moreover, we offer insights into when thesauri and metadata are appropriate to use for such applications.

RELATED WORK

To the best of our knowledge, no other tools currently support comparison search in the cultural heritage domain. We first discuss the state of the art information access tools in cross-collection cultural heritage search sites followed by support for comparison search in other domains.

As a result of many digitization projects, the collections of many museums, archives and libraries are now accessible online. Recently, new aggregated search applications allow users to search directly in multiple collections. Examples include portal sites such as europeana.eu and www.collectiewijzer.nl; and research prototypes such as MultimediaN E-Culture [14] and CultureSampo [8]. Such systems are useful, not only because they enable users to find pieces of information faster, but because of their potential for comparing objects from different collections. Most of these systems include interfaces for common information seeking tasks, namely searching, browsing and exploration [15]. Unfortunately, however, the interfaces of none of the systems mentioned above support comparison tasks directly. Rather, experts are forced to use the standard search and browse interfaces provided.

Related work on comparison search can be found, however, in other domains. For e-commerce applications, analysis of requirements for supporting comparison search has been well covered in the literature [5, 10, 11, 17]. One important requirement of a product comparison search interface is to allow users to extensively search and browse objects before comparing [5, 17]. Selection is also another crucial requirement and can be done in different ways. For example, a study on electronic catalogs [17] emphasizes the importance of allowing users to conduct incremental object selection in comparison search. Another example is to use interactive object filtering based on the available properties, such as in VOPC [11]. In this interface, the properties are visualized next to each other, providing the user with an overview of all possible properties to select from.

A visualization requirement for comparison search is the capability to present multiple properties [11, 17]. Different visualisations have different characteristics. Some presentations are able to show values of a single property (e.g. Bar charts, Dotplots, Histograms and Spinograms), two properties (e.g. Scatterplot, Mosaicplot), or multiple properties (e.g. Table) [19, 25]. Bar charts are simple and straight for-

ward. All values of a single property can be displayed in ascending or descending order. Scatterplots are mostly used when two important characteristics need to be displayed at once, such as in Gapminder [1] and in [12, 13]. Tables are still the most popular visualisations for comparison search in practice. Tables present information in a simple way: the products are presented all in one column, while the product properties are presented in a row (or vice versa) [16, 18]. This type of presentation allows a user to clearly see the values for multiple properties for multiple products at the same time.

The effects of different visualisations on user performance have also been studied, e.g. in [5, 10]. The experiment in [10] shows that table-like interfaces help users solve problems faster, while a scatterplot is better at guiding users to find correct answers. Callahan et al. [5] show that an interactive table (InfoZoom) helps users compare object properties faster than a hierarchical table. The hierarchical table interface, however, was found to be more pleasant to use. The study also suggests that the user's performance, while using a comparison search interface, depends on the type of task, the context and the ability of the user to translate the given problem while working with the system [5]. Thus, different domains and tasks might have different comparison search requirements.

PRELIMINARY STUDY: UNDERSTANDING COMPARISON SEARCH IN THE CULTURAL HERITAGE DOMAIN

We conducted the preliminary study with two goals in mind. First, to identify problems that experts face when they conduct comparison searches. Second, to derive realistic use cases about comparison search tasks that cultural heritage experts carry out during their work.

Setup

We carried out one to two hours semi-structured interviews that took place at the participant's working environment. Each interview consisted of several parts, starting with an introduction explaining the study and general demographic questions. We then asked questions related to accessing multiple cultural heritage sources and how to compare results coming from these sources. Next, we showed sketches of a hypothetical comparison search prototype. These were displayed on the computer to animate interactions with the interface. Finally, the participants were given the opportunity to ask any questions or address concerns about the sketches. In total, seven cultural heritage experts from three different institutions took part in the interviews (see Table 1). The participants' average age was 39 years. They had diverse roles: 2 researchers, 3 curators, 1 art historian and 1 consultant. Most participants had senior positions and had a good overview of the different expert roles within their organization. We hoped that our participants would thus be able to provide insights into the work of their colleagues as well as into their own. All interviews were voice recorded for documentation.

Results

We divide the participants' comments into two themes: the comparison search task use cases that experts conduct for

Table 1. Preliminary Study: Participants demography (total: 7 people)

Age:	35-42 years old ($M=39.3$, $SD=2.8$)
Gender:	1 male, 6 female
Affiliation	CH institution(6), museum(1),
Expert role:	researcher(2), curator(3) art historian(1), consultant(1)

their work, and challenges that experts face in conducting these tasks.

Comparison search use cases

In the interview, we asked participants for instances of comparison search tasks that they, or people in their community, would conduct as part of their daily work. To stimulate ideas, we showed them mock-ups of a comparison search tool. Based on the comparison search demonstrated in the mock-up, participants described several use cases along with the roles of those involved. These use cases can be either qualitative or quantitative comparisons.

Learning about collections — As a part of their education program in art history and museology, students are required to familiarize themselves with the variety of museum collections. Currently, this can only be done by browsing through different museum websites individually. “*Students study sculptures from different museums to get a first impression about what different museums have related to sculptures.*” [P6]

Planning exhibitions — Whenever a curator needs to prepare for an exhibition, s/he needs to find and collect different artwork candidates. This is followed by a selection process where the curator compares and judges each artwork to find the most suitable ones to be displayed in the exhibition. “*I’m preparing an exhibition on [a painter], and need to make a selection and then from the selected items decide which ones should be finally picked for the exhibition.*” [P4]

Museometry¹ — Museum collections change throughout time. Artworks may be loaned, borrowed, sent for restoration or donated. The museum management needs to have periodic quantitative reports of the distribution of artworks to allocate appropriate resources, e.g. for risk management or for expert training courses. “*A museum with diverse collections will benefit more from this interface [mock-up], in particular from the managerial perspective. If I can see, for example, that 80% of my collections are [made of] wood, then I know how many resources to allocate for wood preservation.*” [P2].

Qualitative comparison — Experts often need to conduct qualitative comparisons on other experts’ assumptions, opinions or recommendations. This task may require in-depth cognitive analysis and interpretation. For example, one participant often needs to analyze different point of views “*I would like to compare arguments between experts about a particular topic. First look at a lot of projects and look for best practices. Always choose from internationally recognized studies.*” [P3]

Challenges in conducting comparison search tasks

The challenges mentioned by participants were primarily about: first, *searching* the terms (e.g. name aliases, multiple languages and multiple terms); second, *comparing* objects in multiple sets and multiple properties. Descriptions of these issues are presented below.

Name aliases — The participant may not always specify the correct name when searching because s/he does not know

¹Museometry: research that emphasizes a quantitative approach to answer questions related to different aspects of museum information and its quality.

which variant is used in the collection. For example, location names change with time, e.g. Burma (old name), Myanmar (official name). Artists may also have different name aliases. “*The problem is you don’t always know how to write the artist name that belongs in a specific collection.*” [P2]

Multiple languages — A related problem is when artworks have multilingual annotations. Artworks coming from all parts of the world may be annotated in their vernacular terms or other languages (e.g. Spain, Spanje, España). In order to find these artworks, traditionally, the user needs to perform multiple searches using all possible terms and languages. Not only is this task tedious but also not always obvious for users. “*How could I search for artworks if the language is different?*” [P5]

Multiple terms — There are many potential terms that museums can use to annotate their work. To an outsider, even with some level of domain expertise, guessing which search terms to use is not obvious. One museum curator mentioned that she often needs to help website visitors with their searching. “*Sometimes visitors (of the museum collection website) do not know what to type. For example, to search for an Islamic collection, there are many different words: Islam, Islamic, Moslem, Muslim. Thus sometimes I do the searching for them and send the (search result) link.*” [P6]

Comparing many sets — The tools used do not support the comparison of multiple sets of artworks e.g. comparing the differences between artworks from different museums. For example, one participant researching on museum management usually compares 77 contemporary art museums in the Netherlands at the same time. “*Most of the time I compare more than one museum.*” [P6]. Comparison may be based on various characteristics, such as artworks from different museums or artworks from different artists.

Single and multiple property comparison — Tools do not support comparison search tasks using multiple properties: “*Compare collections of artworks by female artists from before and after the 1960’s.*” [P1]. In another example, the curator wants to highlight different aspects of African art collections in museums in the Netherlands, and wants to compare how different cultures (e.g. Akan, Gurma) predominantly create different artwork types (e.g. mask, painting) and how this changes according to the history of the nation (e.g. pre-colonialism v.s. post-colonialism).

Key findings

The preliminary study provided us insights into comparison search tasks conducted by cultural heritage experts. Our intention was to see where a thesaurus-based comparison search tool can help the experts. Comparison search use cases, such as learning about collections, planning exhibitions and museometry, may require quantitative processing of artwork’s metadata. Since there can be many homonyms and synonyms within the metadata, simple text matching is insufficient. Quantitative computation on the artwork’s metadata taking the thesaurus that provides the metadata terms could yield more accurate results. Therefore, for these use cases, a thesaurus-based comparison search tool maybe useful. For other use cases, such as qualitative comparisons, a thesauri-based comparison tool might not be sufficient. Qualitative comparison requires in-depth analysis and inter-

pretation of information primarily coming from non structured data, such as literature and articles. Thus, for these use cases, metadata and thesauri are insufficient information sources. This is confirmed by participants. For example, when asked about using metadata for comparing different experts' points of view, participants commented: "*Sometimes the information is just too basic.*" [P5], "*This is not specific enough for the job.*" [P2]

Tasks identified were searching and comparing. In order to compare sets results found have to be selected to become part of the comparison tasks. The three main challenges in comparison search: (1) to *search* using the terms that match with the variety in object's metadata, such as name aliases, multiple languages, multiple terms; (2) to *select* sets of objects to compare and (3) to *compare* objects belonging to multiple sets and having multiple properties.

DESIGN REQUIREMENTS

The preliminary study and related work gave us directions on the design requirements for a thesaurus-based comparison search tool for the cultural heritage domain: it should provide features that help users overcome the search, selection and comparison challenges:

Searching artworks — Most of the problems when searching for artworks, such as name aliases, multiple languages and multiple terms, are related to finding terms that match with the artwork's metadata. To help users find the matching terms, a *guided search*, such as an interactive query expansion interface [4], feature should be provided.

Selecting artworks — Selecting artworks is an important part of a comparison search task [11, 17]. Selection is an intermediate step where participants define sets of artworks to be compared against each other. While this activity is enabled in some museum collection websites that cultural heritage experts frequently use [2], the interface and interaction is often unintuitive and sluggish. There are two important requirements for the artwork selection process: first, the selection process has to be easy and convenient; second, to ensure a smooth selection process, the user has to be able to add and to remove any artwork from a selected set.

Comparing artworks — There are two requirements for this activity: first, a comparison search tool for the cultural heritage domain should support comparing multiple sets, where a set can contain many artworks. In the presentation, the differences and similarities between these sets should be clearly distinguishable. Second, it is important to support comparison of single and multiple properties. To focus our research, we concentrate on supporting one property (single property) and two properties (dual property) comparison. For different types of comparison search tasks (single and dual property) a suitable presentation type should be made available [19, 25].

LISA: DESIGN AND IMPLEMENTATION

The LISA application² is part of a suite of tools developed within the MultimediaN E-Culture project³. The project con-

²The LISA prototype is accessible at: <http://e-culture.multimedien.nl/lisa/compsearch>.

³<http://e-culture.multimedien.nl/>

centrates on providing intelligent access to distributed and heterogeneous cultural heritage collections. In the following sections we discuss LISA's technology infrastructure, the datasets and the user interface. We focus our discussion primarily on the interface and interaction. For an extensive discussion on the technological infrastructure, see [14, 24].

Design

Overview — To support the identified activities of comparison search, the LISA interface consists of four areas (see Fig.1 top left): (a) the search area, (g) the search result area, (k)(l) the selection areas, and (j) the comparison area. In the search area, the user incrementally formulates queries, and the results are shown in the search result area below. In the selection area, users can see two sets of artworks. In the comparison area, there are alternative visualizations that can be used to see the characteristics of the selected artworks. We discuss the three supported activities separately. As an example, we compare the self-portraits of Vincent van Gogh with the self-portraits of Rembrandt van Rijn (see Fig.1).

Search — To support search, we use a thesaurus-based guided search that consists of a property filter and autocompletion in the interface. The property filter is a pull down menu that shows all possible artwork properties, such as artwork creator, creation date or material, that can be selected (see Fig.1b). When a user types a keyword, the autocompletion interface will show suggestions of terms used by the museum collections. An alphabetical ordering is used for our autocompletion suggestions as we found it to be the most effective ordering for loosely structured thesauri [4]. The guided search takes the form of `property:value` pairs. Users can complete a search by selecting an autocompletion suggestion (Fig.1d). To assure full flexibility, the user can add and remove as many `property:value` pairs as s/he wishes. In the example of Fig. 1, to search for self-portraits of Rembrandt van Rijn, the user specified two guided searches: to search for all artworks having a `Creator:Rembrandt van Rijn` *and* `subject:zelfportret` (English: self-portrait) (see Fig 1c), and to search for all artworks having a `Creator:Vincent van Gogh` *and* `subject:zelfportret`.

Because related terms from different thesauri are linked, the system can provide a match even though the artworks' metadata are different as long as they are semantically equivalent. For example, there are 31 name aliases for 'Rembrandt van Rijn' in ULAN⁴, e.g. 'Rembrandt van Ryn' or 'Rembrandt van Rhijn'. If a user specifies any of these alternative names, LISA will be able to retrieve the same artworks. Similarly, it is possible to thoroughly search with different geographical, art and architecture, and iconographic name aliases. For a list of all linked thesauri supported by LISA, refer to Table 2. For more information about designing and configuring a thesaurus-based autocompletion see [4]⁵.

Selection — After the user is satisfied with the search results, s/he needs to select the artworks to compare. The

⁴Union List of Artist Names thesaurus, http://www.getty.edu/research/conducting_research/vocabularies/ulan/

⁵In this implementation, only syntactic matches are shown in the suggestions. This is, however, configurable to also suggest semantic matches, such as broader, narrower and related terms.

system allows the user to add multiple artworks in either set. In the current implementation, we only support comparison between two sets ⁶. An artwork can be placed in any of the available sets (Set A or Set B). There are two easy ways to add an artwork to a selection: first, by dragging and dropping an artwork thumbnail from the search result panel to the selection panel (Fig. 1k,l); second, a bulk selection of all search results can be made by clicking the *Set A* or *Set B* button (Fig. 1f). To allow fine tuning of selections, adding or removing an artwork from the set is made possible. The search and selection process are typically done sequentially. For example, first the user searches for all self-portraits of Van Gogh (`creator: Vincent van Gogh, subject: zelfportrait`), places them on the Set A selection area (Fig. 1k). Afterwards s/he makes a second search of all self-portraits of Rembrandt (`creator: Rembrandt van Rijn, subject: zelfportrait`) and places the results in the Set B selection area (see Fig. 1l).

Comparison with visualizations — LISA currently supports single property comparison and dual property comparison. We choose the Bar chart (Fig. 1j3) for single property comparison and the Scatterplot (Fig. 1j1) for dual property comparison because these presentations are the most common from a variety of visualizations specified in [19, 25]. Additionally, we also implemented a Table visualization because this type of presentation is the default presentation for most comparison search applications (see Fig. 1j2). Whenever an object is placed in the selection area (Set A or Set B), the visualization area is updated. In all visualizations, we use color codes to indicate which sets the elements belong to (either Set A or Set B). Fig. 1j3 shows a Bar chart representation of artworks from both sets. The bar chart highlights the comparison between the two sets with respect to the chosen property: `material`. The x-axis represents the artworks organized by the selected property: `material` in alphabetical order. The y-axis represents the number of artworks. The figure shows the values of Set A and Set B next to each other. The property pull-down menu (Fig. 1o) shows all available properties for which the objects can be organized, e.g. by dimension height, date, material or depicted subject. The scatterplot presentation (Fig. 1j1) shows comparison of sets with respect to the dual property selection (Fig. 1i) i.e. `material` (y-axis) and `date` (x-axis). This presentation highlights the differences between the two sets with respect to the creation time and what materials they are made of. The table visualization (Fig. 1j2) shows all artworks from Set A. To view all properties from Set B, the user needs to select the tab (Fig. 1m). It is possible to explore the information space by two means: a) alternating between different properties by selecting an item at the property pull down menu (Fig. 1o), or b) visual exploration either by zooming, panning or scrolling. With the different visualizations, the user can flip through different properties to examine multiple collections simultaneously to gather quick insights about vast collections, which is extremely difficult with current tools.

⁶This is extendable as the current application design takes into account the future addition of more sets, however, since computations are carried out client-side, for more than 1,000 objects the interface becomes slow.

Table 2. Thesauri and collections used in LISA
source (thesaurus coverage) size

Collection:		
RKD Archive	82.781	objects
Thesaurus:		
RKD thesaurus (RKD)	11.995	terms
TGN (geographical)	89.000	terms
ULAN (artist)	13.000	people
AAT (art and architecture)	31.000	terms
IconClass (iconographic)	24.331	terms

Implementation

Infrastructure — The LISA application is developed on top of *ClioPatria*, a web application platform for search and annotation across heterogeneous collections. For detailed information on the web server infrastructure and the search strategies across heterogeneous collections, see [14, 24]. Communication between the client and the server is done via requests to the system's HTTP API. Information is sent back from the server in JSON. The implementation of the interface uses (X)HTML, CSS, Javascript and Flash. It is tested on the Firefox 3.0.10 browser. The client side visualization widgets use an extension of the Yahoo User Interface widget (YUI v. 2.7.0) and amChart v. 1.6.5 ⁷.

Dataset — To enable comparison search with LISA, the server needs to host common thesauri, namely IconClass⁸, the Getty Art & Architecture Thesaurus (AAT), the Getty Union List of Artist Names (ULAN) and the Getty Thesaurus of Geographical Names (TGN) ⁹, as well as collection specific thesauri, such as thesauri from RKD. Table 2 shows the size of collections and thesauri currently used by the application. Collections and thesauri data were converted to an RDF/SKOS representation. To allow information access across collections, specific thesauri are aligned with the common ones. For example, artists' names in the RKD thesauri are linked to artists' names in ULAN. Materials terms in the RKD thesauri are linked with concepts in AAT. Detailed information on the conversion and alignment methods of cultural heritage sources used can be found in [20, 22].

EVALUATION STUDY: THESAURUS-BASED COMPARISON SEARCH INTERFACE EVALUATION

The goal of this study was to evaluate LISA. In particular, we focus on evaluating how well the search, selection and comparison features support experts' comparison search tasks. As a baseline, we use the RKDimages website [2], a popular online cultural heritage archive that contains descriptions, metadata and images of Dutch and Flemish artworks from the 14th-19th century. The RKDimage website contain a comprehensive coverage of different artworks and is widely used as reference. For the purpose of the evaluation, participants can access the same information with LISA as well as with RKDimages website. The research questions for this evaluation are:

⁷<http://www.amChart.com>

⁸<http://www.iconclass.nl>

⁹http://www.getty.edu/research/conducting_research/vocabularies/

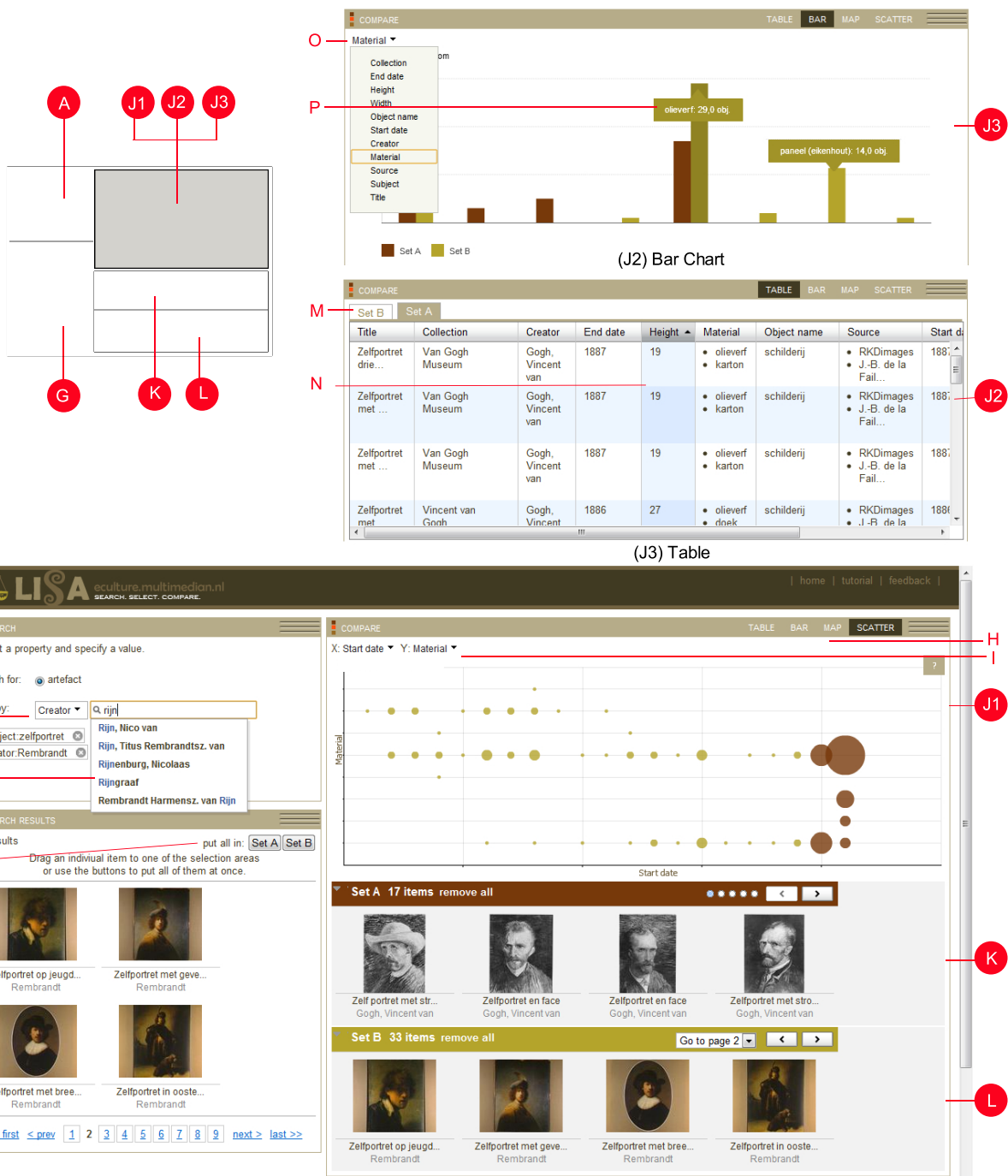


Figure 1. The LISA interface (a) search area, (k) selection area Set A, (l) selection area Set B, (j) comparison area: (j1) Scatterplot (j2) Table (j3) Bar chart
Features: (b) property filter, (c) guided search, (d) autocompletion suggestions, (e) number of search results, (f) selection shortcut, (g) search results, (h) visualization selection options, (i) Scatterplot 2 properties selection,
 (j1) The scatterplot shows the distribution of Van Gogh's self portraits and Rembrandt's self portraits with respect to the date and material used. The visualization shows that Rembrandt had consistently painted a small number of self portraits distributed throughout many years using different kinds of materials (e.g. canvas, Oak panel, Mahogany panel). Van Gogh, however, made many self-portraits between 1886 and 1887. In 1887 alone, he made about 11 oilpaint (olieverf) self-portraits.
 (j2) The table shows all values of Van Gogh's self-portraits. The smallest height of a van Gogh's self-portrait is 19 cm.
 (j3) The bar chart shows the distribution of materials used for Set A and Set B. 29 of Rembrandt's self-portraits are made out of oilpaint (olieverf) and 14 are made out of Oak (Paneel Eikenhout).

Table 3. Evaluation Study: Participants demography (total: 12 people)

Age:	21-60 years old ($M=39.6$, $SD=12.1$)
Gender:	4 male, 8 female
Affiliation	CH institution(6), museum(2), Art historical archive(3), university(1)
Expert role:	researcher(4), curator(2), ICT(2) program manager(3), librarian(1)

- Does the LISA tool support searching, selecting and comparing artworks more efficiently than the baseline tool?
- Do users perceive the LISA tool easier to use than the baseline tool?

Setup

The experiments took place at the participant's working place. Each participant was asked to complete comparison search tasks with LISA and RKDimages (within subject design). The experiment had four parts:

- Introduction. Participants were asked demographic questions and for informed consent. Afterwards, they were shown a video demo and were asked to perform trial tasks on the LISA application. As most participants were already regular users of the RKD website, the demo focused on familiarizing the participants with LISA.
- Experiment session. The experiment was divided into two phases. In the first phase, we compared how well the 4 different type of presentations (LISA Table, LISA Bar Chart, LISA Scatterplot, RKDimages) support single property comparison. In the second phase, we compared how well 3 different type of presentations (LISA Table, LISA Scatterplot, RKDimages) support dual property comparison¹⁰. In total, participants were given 14 comparison tasks (2 tasks per interface). At the completion of each task, participants were asked about how they perceived the ease of use of the interface used. At the end of all tasks, participants gave general impressions about LISA and RKDimages.

User recruitment was based on email invitations and open invitations on a cultural heritage online forum. In total there were 12 cultural heritage experts from seven cultural heritage institutions (see Table 3). Our participants conduct searches within collections frequently for their work ($M=4.8$, $SD=1.8$)¹¹ and are fairly familiar with the RKDimages website ($M=3.8$, $SD=2.5$)¹². None of the participants had used the LISA interface prior to this evaluation.

Task

We use two independent variables as dimensions that reflects the complexity of comparison search.

- Number of artworks. We have two conditions: comparing *few* objects (1 artworks per set) and comparing *many* artworks (10-15 artworks per set). We expect that the more artworks there are, the harder it is to perform comparison search.

¹⁰The LISA Bar chart was omitted in the second phase because it is unsuitable for dual property comparison task

¹¹Seven point scale, 1: not very often, 7: very often

¹²Seven point scale, 1: very unfamiliar, 7: very familiar

- Number of properties. We have two conditions: *single* property comparison and *dual* property comparison. An example of a comparison search task is as follows:

Use the scatterplot to answer this question: Compare all artworks having the subject depiction "church" from the Stedelijk Museum De Lakenhal with all artworks having the subject depiction "church" from Museum Bredius.

(1) Which artist made the most artworks?

(2) How many of these artworks are paintings and are made after 1612?

Example 1 is a single property comparison task for the property `Artist`. Example 2 is a dual property comparison task for the properties `Object type` and `Date`.

Results

In this section we discuss observations on how well LISA and the RKDimages website support the comparison search tasks conducted by the participants.

Searching and selecting artworks — Prior to the study, we naively assumed that the search, selection and comparison activities are separate and independent. However, during our experiments we observed that the search and selection activities were highly interdependent. Participants conduct a search, make selections and refine the search before being satisfied with the selection result. Only then do participants carry out the more independent comparison activity. This non-linear behavior is also suggested by previous research, such as in [9]. Thus, in our discussions, we describe search and selection activities together and the comparison activity separately.

We analyzed the time performance and the perceived ease of use for searching and selecting artworks. Table 4a. shows the average time it took for participants to complete searching and selecting for (Few and Many) artworks with LISA and RKDimages website. Using the Wilcoxon Sign-rank (WSR) test¹³, we found that participants were about two times slower when searching and selecting many artworks using RKDimages ($Mdn=2.75$ min) than searching and selection few artworks ($Mdn=1.12$ min) $z=-2.04$, $p<.05$, $r=-.42$. This confirms our expectation that the more artworks there are, the more time it will take to search them using the baseline tool. When participants use LISA for searching and selecting artworks, however, the results were different. We did not find a significant difference when participants were searching and selecting few artworks ($Mdn=1.43$ min) compared with many artworks ($Mdn=1.15$ min). Moreover, when searching and selecting many artworks, RKDimages ($Mdn=2.75$ min) is significantly slower than LISA ($Mdn=1.15$ min) $z=-2.43$, $p<.05$, $r=-.49$. Thus, we conclude that LISA is more time efficient than RKDimages for searching and selecting many artworks.

Participants also rated the perceived ease of use of LISA and RKDimages for searching and selecting artworks (see Table 4b,c). RKDimages is significantly easier to use when

¹³ Nonparametric statistics is used throughout the study as not every data meet parametric assumptions.

Table 4. Searching and Selecting artworks with LISA and RKDimages

	LISA	RKDimages
a. Search and select avg. time in min (SD)		
1. FEW	1.38(.41)	1.58(1.19)
2. MANY	1.37(.51)	3.33(1.90)
b. Search - ease of use score (SD)		
1. FEW	5.75(1.22)	5.58(1.38)
2. MANY	6.08(1.08)	4.67(1.78)
c. Select - ease of use score (SD)*		
1. FEW	5.92(1.16)	4.50(1.68)
2. MANY	6.00(0.95)	3.58(1.68)

* 7-Likert point scale, score 1:very difficult, 7:very easy

Few: few objects comparison (1 item/set)

Many: many objects comparison (10-15 items/set)

searching for few artworks ($Mdn=6.0$) than when searching for many artworks ($Mdn=5.0$) $z=-2.46$, $p<.05$, $r=-.50$. When searching with LISA, however, this difference was not significant. Searching for many artworks ($Mdn=6.0$) is perceived as easy as searching for few artworks ($Mdn=6.0$). Participants also think that RKDimages ($Mdn=5.0$) is harder to use than LISA ($Mdn=6.0$) when they need to search for many artworks $z=-2.38$, $p<.05$, $r=-.48$. Thus, we conclude that LISA is easier to use than RKDimages, especially for searching for many artworks.

We asked the same question for selecting artworks and found congruent results. We found LISA significantly easier to use than RKDimages when selecting few artworks ($z=-2.06$, $p<.05$, $r=-.42$) as well as many artworks ($z=-2.69$, $p<.05$, $r=-.55$). Based on this, we conclude that LISA is easier to use than RKDimages with respect to selecting few as well as many artworks.

For searching and selecting artworks, we found LISA more time efficient and easier to use than RKDimages. We also found that, unlike the time efficiency with the baseline tool, time efficiency with LISA does not suffer much as there are more artworks to search and select. Moreover, overall, participants perceived LISA easier to use than RKDimages for searching and selecting artworks.

Comparing artworks — We use Friedman Analysis of Variance by Ranks (FAVR) to examine if there were any differences in time performance between the different presentations types for single property comparison and dual property comparison tasks (see Table 5).

• **Single property comparison search tasks**

For single property comparison, we compared participants' time efficiency using four different presentations: LISA-Table, LISA-Bar chart, LISA-Scatterplot and RKDimages. Using the FAVR test, we found no significant difference for the four different presentations. This applies for a single property comparison task for few artworks ($\chi^2(3)=3.7$, $p>.05$) as well as for many artworks ($\chi^2(3)=3.0$, $p>.05$) We conclude that all four presentations perform equally with respect to the time spent to conduct single property comparison.

• **Dual property comparison search tasks**

Table 5. Comparing artworks with LISA and RKDimages

a. Single property comparison avg. time in min. (SD)				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	0.91(.75)	0.94(.34)	1.32(.66)	0.83(0.24)
2. MANY	1.37(.55)	1.19(.36)	1.70(.76)	2.13(1.61)
b. Dual property comparison avg. time in min. (SD)				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	1.14(.71)	-	1.28(.63)	1.06(.40)
2. MANY	2.28(.97)	-	1.39(.65)	2.99(.99)
c. Compare - ease of use score (SD) *				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	4.92(1.68)	5.42(1.08)	4.50(1.78)	2.83(1.47)
2. MANY	5.08(1.38)	4.75(1.54)	4.75(1.76)	2.17(1.47)

* 7-Likert point scale, score 1:very difficult, 7:very easy

In dual property comparisons, we found no significant difference between the three presentations for few artworks comparisons ($\chi^2(2)=2.2$, $p>.05$). However, there is a significant difference between three different presentations (i.e. LISA-Table, LISA-Scatterplot and RKDimages) for many artworks ($\chi^2(2)=9.5$, $p<.05$). We found significance of time performance between all three presentations. LISA-scatterplot being the fastest ($Mdn=1.26$ min), followed by LISA-table ($Mdn=2.08$ min), and RKD images ($Mdn=3.02$ min) the slowest¹⁴.

We saw clearer trends for the perceived ease of use scores (see Table 5c). We found significant differences for comparison search tasks in few artworks ($\chi^2(3)=13.07$, $p<.05$) as well as many artworks ($\chi^2(3)=17.49$, $p<.05$). The WSR post-hoc test confirms that all LISA visualizations are perceived easier to use than RKDimages for comparison.

We conclude that comparing with LISA is more time efficient than with RKDimages mainly for many artworks dual property comparison search task. However, participants perceive LISA easier to use than RKDimages for few as well as many artworks comparison.

DISCUSSION

The thesaurus-based comparison search tool builds on two important components: the interface and the data. We discuss challenges and improvements with respect to these.

Interface

Searching and selecting artworks — The evaluation study showed that searching and selecting artworks with LISA gives better time performance than with RKDimages. This is mainly because of the ease of use of the thesaurus-based guided search and the selection interface. The thesaurus-based auto-completion enable users to quickly find the correct term to find artworks. This aligns with previous research on the use of thesaurus-based auto-completion for term search [4]. To increase confidence levels when selecting terms from the au-

¹⁴ WSR post-hoc tests shows LISA Scatterplot - LISA Table ($z=-2.12$, $p<.05$, $r=-.43$), RKDimages - LISA Table ($z=-1.96$, $p<.05$, $r=-.40$), RKDimages - LISA Scatterplot ($z=-2.83$, $p<<.05$, $r=-.58$),

to completion suggestions, improvements, such as adding extra information about the terms, e.g. by showing the thesauri hierarchy or descriptions of the terms, could be made [6].

Comparing artworks — We found that, with respect to time spent, LISA shows little improvement on RKDimages for comparison activities. There can be several explanations for this. Our participants are regular RKDimage users, thus they are more experienced in using this application for comparison. Most of them, however, are not used to handling graphs and charts. We observed that participants need to spend time to become familiar with the tool. We acknowledge that we cannot fully eliminate learning effects in a one time evaluation study of a complex tool such as LISA. Many of the participants thought they would be able to handle LISA better once they were accustomed to it. *“I think you have to get used to the system, like to any new system”* [P1], *“This is a new way to present and interact with (museum) collections.”* [P14].

Even though LISA did not significantly improve the speed of comparison, participants clearly favor LISA above RKDimages with respect to ease of use. They see the practical benefit of having aggregated results presented automatically rather than computed manually. Before trusting the results, however, they need to understand how the thesaurus-based aggregation works to produce the graphs and charts of the presentations. Participants also appreciated the different presentations as they provide more ways to analyze the same data. *“I think there are different learning and reading styles, so it is useful to have these variations”* [P3].

Additional features — Sometimes experts need to go back and re-examine previous comparison search tasks. Experts may also want to save the visualization results and include them in a report, or may need to inspect the visualization in detail. Our experts listed features such as bookmarking, search history and ‘save as’ as additional functionalities that LISA should have. Participants also mentioned two important visualization improvements. First, the ability to enlarge the visualization size on demand. This feature is especially useful when dealing with many artworks. Second, more interactivity with the visualization, such as being able to trace back from the visualization to the original artworks.

Data

Based on what we have learned during the LISA implementation and the evaluation study, we identify characteristics of the collection metadata that developers need to be aware of when developing thesaurus-based aggregator services.

Semantic aggregation — Thesaurus-based comparison search should take into account semantic aggregation where narrower/broader relationships exist between terms. For example, to be able to answer the question “How many artworks are paintings?” correctly, the system needs to quantitatively aggregate not only all artworks having `object type: painting`, but also all artworks annotated by the narrower concepts of painting, such as `object type: aquarel`, since Aquarel is a type of a painting.

Inconsistent data — Museum collection metadata may be inconsistent, for example in measurement units. Artwork dimensions, such as height and width, can be specified in

different units, e.g. feet, cm or mm. Prior to an aggregation process, metadata needs to be cleaned.

Incomplete metadata — In reality, museum metadata is not always complete. Parts of the collection may have insufficient or missing values. A quantitative aggregation on these data will generate false results. One solution is to check and improve the quality of data automatically as suggested in [23].

Estimated data — In some cases, the metadata contains an estimated value, e.g., the creation date of an artwork is simply unknown. The museum is able to supply only an estimate, e.g. “before AD 400” or “between 600-700”. Providing accurate aggregation results based on these estimated data is not possible.

Quality of data alignment — The accuracy of information presented in a thesaurus-based aggregated system also depends on the quality of the data alignment, i.e. linking metadata to terms from individual thesauri and linking terms among different thesauri. Methods for vocabulary alignment are still at a preliminary stage of development, e.g. [21], making it difficult to predict when automatic methods will be of sufficient quality for our experts’ needs.

The biggest opportunity for LISA lies in the area of comparison across different collections. At the moment, such tasks are carried out by accessing the different collections individually, and then integrating the data manually. While there is room for improvement by being able to automatically aggregate over multiple collections, the number of errors in the results will also grow when performing computations across collections because of the different schemata and thesauri used to describe the artworks. When dealing with real museum datasets, we believe that the answer needs to come from making the computations more transparent in the interface, allowing the user to examine the data and how the computation is being performed. Thus, if there are errors in the results, the user should be able to trace them, make corrections and even correct the underlying data.

CONCLUSIONS AND FUTURE WORK

We conducted a user-centered design study on a comparison search application for the cultural heritage domain. In a preliminary study, we identified various comparison search use cases, such as learning about collections, planning exhibitions, museometry and qualitative comparison, and identified the challenges users face while performing comparison search, such as searching and selecting terms, and comparisons involving multiple sets and properties. In our evaluation study, we found that our comparison search tool can help users, in particular for efficient searching for terms and selecting artworks. In general, participants perceived the comparison search tool as easier to use than the baseline tool with respect to searching, selecting and comparing artworks. Finally, based on our implementation experience and evaluation study, we identify future improvements to the interface, namely, supporting interactivity visualizations, improving the autocompletion and providing bookmarking and search history functionality. For the data issues, we would be able to extend LISA’s functionality to support semantic aggregation. For the other data issues we are dependent on oth-

ers making the data more reliable. Making the issues more transparent through the LISA interface would be one way of tackling this.

ACKNOWLEDGEMENTS

We thank the cultural heritage experts from Centraal Museum Utrecht, Digital Heritage Netherland (DEN), Efgoed Nederlands, Netherlands Collection Institute (ICN), Publiek Archief Eemland, Netherlands Institute for Art History (RKD), Rijksmuseum Amsterdam, Tropenmuseum, and University of Amsterdam. We thank Hyowon Lee for his feedback during the interface design. This research was supported by the MultimediaN project through the BSIK program of the Dutch Government and by the European Commission contract FP6-027026, KSpace.

REFERENCES

1. <http://www.gapminder.org/>.
2. <http://english.rkd.nl/databases>.
3. A. Amin, L. Hardman, J. van Ossenbruggen, and A. van Nispen. Understanding cultural heritage experts' information seeking tasks. In *JCDL '08*, pp 39–47, NY, USA, 2008. ACM.
4. A. Amin, M. Hildebrand, J. van Ossenbruggen, V. Evers, and L. Hardman. Organizing Suggestions In Autocompletion Interfaces. In *ECIR '09*, pp 521–529, 2009.
5. E. Callahan and J. Koenemann. A comparative usability evaluation of user interfaces for online product catalog. In *EC '00*, pp 197–206, NY, USA, 2000. ACM.
6. M. Hildebrand, J. van Ossenbruggen, and L. Hardman. Supporting subject matter annotation using heterogeneous thesauri: A user study in web data reuse. *IJHCS*, 67:887–902, Oct 2009.
7. M. Hildebrand, J. R. van Ossenbruggen, A. K. Amin, L. Aroyo, J. Wielemaker, and L. Hardman. The Design Space Of A Configurable Autocompletion Component. Technical Report INS-E0708, CWI, 2007.
8. E. Hyvonen, et. al. Culturesampo – finnish cultural heritage collections on the semantic web 2.0. In *Proceedings of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures (DH-JAC-2009)*, Ritsumeikan, Kyoto, Japan, March 2009.
9. A. Kerne and S. M. Smith. The information discovery framework. In *DIS '04*, pp 357–360, NY, USA, 2004. ACM.
10. A. Kobsa. An empirical comparison of three commercial information visualization systems. In *INFOVIS '01*, pp. 123, Washington, DC, USA, 2001. IEEE Computer Society.
11. J. Lee, H. S. Lee, and P. Wang. An interactive visual interface for online product catalogs. *Electronic Commerce Research*, 4(4):335–358, 2004.
12. S. Perugini, K. McDevitt, R. Richardson, M. A. Pérez-Qui R. Shen, N. Ramakrishnan, C. Williams, and E. A. Fox. Enhancing usability in citidel: multimodal, multilingual, and interactive visualization interfaces. In *JCDL '04*, pp 315–324, NY, USA, 2004. ACM.
13. C. Plaisant, J. Rose, B. Yu, L. Auvil, M. G. Kirschenbaum, M. N. Smith, T. Clement, and G. Lord. Exploring erotics in emily dickinson's correspondence with text mining and visual interfaces. In *JCDL '06*, pp 141–150, NY, USA, 2006. ACM.
14. G. Schreiber, et. al. Semantic annotation and search of cultural-heritage collections: The multimedian e-culture demonstrator. *J. Web Sem.*, 6(4):243–249, 2008.
15. R. Shen, N. S. Vemuri, W. Fan, R. da S. Torres, and E. A. Fox. Exploring digital libraries: integrating browsing, searching, and visualization. In *JCDL '06*, pp 1–10, NY, USA, 2006. ACM.
16. M. Spence, C. Beilken, and T. Berlage. Focus: the interactive table for product comparison and selection. In *UIST '96*, pp 41–50, NY, USA, 1996. ACM.
17. P. Steiger and M. Stolze. Effective product selection in electronic catalogs. In *CHI '97*, pp. 291–292, NY, USA, 1997. ACM.
18. T. Tenev and R. Rao. Managing multiple focal levels in table lens. In *INFOVIS '97*, pp. 59, Washington, DC, USA, 1997. IEEE Computer Society.
19. M. Theus and S. Urbanek. *Interactive Graphics for Data Analysis*. Computer Science and Data Analysis, 2009.
20. A. Tordai, B. Omelayenko, and G. Schreiber. Semantic excavation of the city of books. In *SAKM*, 2007.
21. A. Tordai, J. van Ossenbruggen, and G. Schreiber. Combining vocabulary alignment techniques. In *K-CAP '09*, pp 25–32, NY, USA, 2009. ACM.
22. M. van Assem, V. Malaisé, A. Miles, and G. Schreiber. A method to convert thesauri to skos. *The Semantic Web: Research and Applications*, pages 95–109, 2006.
23. A. van den Bosch, M. van Erp, and C. Sporleder. Making a clean sweep of cultural heritage. *IEEE Intelligent Systems*, 24(2):54–63, 2009.
24. J. Wielemaker, M. Hildebrand, J. Ossenbruggen, and G. Schreiber. Thesaurus-based search in large heterogeneous collections. In *ISWC '08*, pp. 695–708, Berlin, Heidelberg, 2008. Springer-Verlag.
25. F. W. Young, P. M. Valero-Mora, and M. Friendly. *Visual Statistics: Seeing Data with Dynamic Interactive Graphics*. Wiley Series, 2006.