

A Subjunctive Exploratory Search Interface to Support Media Studies Researchers

Marc Bron
ISLA, University of Amsterdam
Science Park 904
1098XH Amsterdam
m.m.bron@uva.nl

Jasmijn van Gorp
TViT, Utrecht University
Muntstraat 2a
3514EV Utrecht
j.vangorp@uu.nl

Frank Nack
ISLA, University of Amsterdam
Science Park 904
1098XH Amsterdam
nack@uva.nl

Maarten de Rijke
ISLA, University of Amsterdam
Science Park 904
1098XH Amsterdam
derijke@uva.nl

Andrei Vishneuski
ISLA, University of Amsterdam
Science Park 904
1098XH Amsterdam
a.vishneuski@uva.nl

Sonja de Leeuw
TViT, Utrecht University
Muntstraat 2a
3514EV Utrecht
s.deleeuw@uu.nl

ABSTRACT

Media studies concerns the production, interpretation and reception of various types of media. Today's continuous production and storage of media is changing the way these researchers work and requires the development of new search models and tools. We investigate the research cycle of media studies researchers and find that it is an iterative process consisting of several search processes in which data is gathered and the research question is refined. Changes in the research question, however, trigger new data gathering processes. Based on these outcomes we propose a subjunctive exploratory search interface to support media studies researchers in refining their research question in an earlier stage of their research. To assess the subjunctive interface we conduct a user study and compare to a traditional exploratory search interface. We find that with the subjunctive interface users explore more diverse topics than with the standard interface and that users formulate more specific research questions. Although the subjunctive interface is more complex this does not decrease its usability. These findings suggest that the subjunctive interface supports media studies researchers. The advantage of a subjunctive interface for exploration suggests a new direction for the development of exploratory search systems.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Search process; H.5.2 [User Interfaces]: Evaluation/methodology

General Terms

Experimentation, Human Factors

Keywords

Exploratory search, subjunctive interface, humanities

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.

SIGIR '12, August 12–16, 2012, Portland, Oregon, USA.
Copyright 2012 ACM 978-1-4503-1472-5/12/08 ...\$10.00.

1. INTRODUCTION

Media Studies can be situated both in the humanities as well as the social sciences and concerns the study of production, content and/or reception of various types of media, e.g., social media, film, and television [22]. Today's continuous production and storage of all types of media is changing the way humanities researchers work and the questions they seek to answer [3, 9, 26]. As a result, there is an increasing demand for tools to support exploration and analysis of this data [31, 32]. Tools to support humanities researchers exist but they are often not effective or remain unused [9, 33]. One reason is that, generally, tools are developed without looking at the context and complexity of the task [16]. Moreover, the level of detail at which available models of the research cycle describe search behavior is insufficient to design effective tools [31].

In this paper we start out by investigating the research cycle of media studies researchers. Through an analysis of a series of interviews about the research habits of media studies researchers we arrive at a model of the media studies research cycle. We compare our model to other models of the humanities research cycle, and find that it is consistent with existing models [4, 8] but that it is more detailed in identifying sequences of search processes and their influence on the research question.

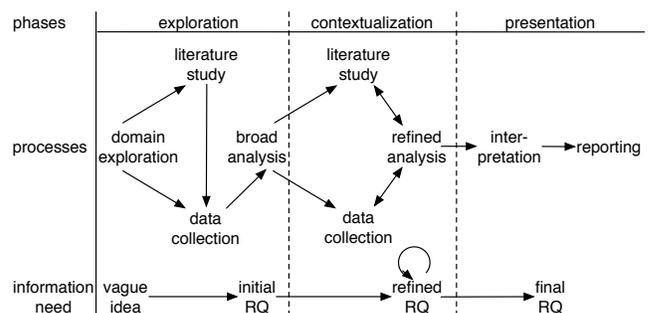


Figure 1: Overview of the phases in the media studies research cycle with associated search processes and changes in the research question (RQ). Arrows indicate possible sequences.

Figure 1 shows the phases in the media studies research cycle with the associated search processes and changes in the research question. We find that the research cycle of media researchers is an iterative process and that the research question is refined in each step.

Take for example the following initial research question: “How are eastern European migrants represented on television,” and its final form: “How are children of USSR refugees represented in television programs for children from 1980 to 2000.” Every change in the research question, however, triggers new data gathering processes. Changing the research question comes at a cost as new data has to be collected, organized and analyzed. We observe that factors influencing the research question are discovering alternative views on a topic and discovering trends in the data. These observations suggest that an interface that supports these types of discoveries in an early stage of the research cycle will reduce the number of iterations necessary to arrive at a final research question.

In this paper we focus on developing and assessing a search interface to support media studies researchers in refining their research question in the *exploration* phase of the research cycle. Requirements for such an interface are that it supports (i) exploring multiple views on a topic and (ii) discovering patterns in the data.

A number of exploratory search tools exist, that support various ways to explore collections, e.g., through filtering by facets, or relevance feedback, see [11, 28, 36] for overviews of systems and their capabilities. As work on exploratory search systems focuses on supporting exploration in a general setting, few have considered how exploratory search systems can support researchers in discovering alternative views and trends in the data during exploration. To support users in complex search tasks [16], investigating multiple aspects of a topic in a subjunctive interface was shown to reduce task complexity and task completion time [23, 34]. In the humanities, standard practices to discover patterns in data are organizing and comparing [25, 27, 32].

We propose to extend the traditional exploratory search system design in two ways: we incorporate two side-by-side versions of an exploratory search interface in a single interface and second, we add visualizations in which the characteristics of the result sets are shown and can be compared. An interface that incorporates multiple instances of the same search tool in a single interface is referred to as a subjunctive interface [23]. Given this subjunctive exploratory search interface we aim to answer the following research questions: (i) does the subjunctive exploratory search interface better support media studies researchers in a complex exploratory search task than a standard exploratory search interface; (ii) does the subjunctive exploratory search interface better support media studies researchers in refining a research question than a standard exploratory search interface; (iii) does the increase in complexity in terms of additional features affect the usability of the subjunctive interface as compared to a standard exploratory search interface?

In §2 we investigate the media studies research cycle. In §3 we describe the subjunctive exploratory search interface and in §4 the experimental design. In §5 we present the results of assessing the subjunctive interface, and provide a discussion in §6. Related work is discussed in §7 and we conclude in §8.

2. MEDIA STUDIES’ RESEARCH CYCLE

To develop an exploratory search tool to support media studies researchers, it is necessary to gain insight into the user’s information behavior and context of the task [15, 38]. Existing models of the humanities research cycle or scholarly search behavior [2, 10, 18, 38] are either too general for building tools [31], do not consider the context of increasing amounts of digital information [8], or focus on a different aspect of the research cycle [4].

We take a detailed look at the research habits of media studies researchers. We take a bottom-up approach and through a series of interviews identify changes in the research questions and search processes of media studies researchers during their research. We

construct a model of their research cycle and link this to existing models of the humanities research cycle and information behavior.

2.1 Method

We interviewed seven media studies researchers from 3 different institutes with varying levels of experience: 1 PhD student, 4 post-doctoral researchers, 1 assistant-professor, and 1 full professor. Two interviewers conducted the interviews: a media studies researcher experienced in conducting interviews and a computer scientist with some interview experience. The interviewers’ backgrounds supplement each other and limit the possibility of misunderstandings during the interview. The subjects were recruited during visits to the 3 institutes and participation was voluntary.

The interview was conducted in a semi-structured style and consisted of three parts: (i) identification of a recent research project; (ii) open questions about search processes and research questions during the project; and (iii) an interactive part in which subjects wrote down the search processes on index cards and ordered them chronologically. This methodology is based on [8]. Interviews lasted about 30 minutes, were tape-recorded and later transcribed.

In our analysis of the interviews we used the following procedure: first we align the search process sequences produced by each of the interviewees by grouping cards with processes that exactly match in the same category. Each card that does not match is placed in a separate category. In the next step we apply coding, a qualitative categorization strategy [30], to the transcripts to identify categories of search processes in the interviews. Categories from the interviews that match with existing categories in the sequences are merged while newly identified categories are added to the sequences. This is an iterative process which is repeated until a stable set of categories emerges. The result is a sequence of categories showing the search processes over time, see Figure 1. Note that when using quotes, square brackets [...] indicate modifications to the original quote to improve understanding or to protect the anonymity of the subject. For identification purposes interviewees are assigned a number, i.e. I1 to I7.

2.2 Results

Three categories (phases) emerged from the analysis: exploration, contextualization, and presentation. Each phase consists of a number of search processes and affects the research question, see Figure 1. Below we describe each category in detail.

Exploration. In general a research project starts with exploration. All but one of the interviewees indicated that they started out with a broad question. One already started with specific research questions as these were defined in an earlier project (I6). With a broad question in mind researchers start an exploratory search for literature, relevant data or both. The goal of this phase is to arrive at an initial research question and to collect a broad sample of data. One researcher noted I1: “for me it starts with developing the research questions and data collection. This happens in parallel, so the question changes by the material you see” and another noted I2: “the data influences the research question, because the data is not available or because you start to see, oh this is so naïve.”

Once a selection of relevant data has been made the initial analysis starts by organizing and structuring the collected material. This is a time consuming process in which material is watched, transcribed, and archived [27]. The goal is to get an overview and to formulate an initial research question, leading to the next phase.

Contextualization. The contextualization phase is the most iterative part of the research cycle. The arrows going back and forth between analysis and data collection, and analysis and literature study in Figure 1, indicate that multiple iterations may be performed. In

this phase a more focused data collection starts to place the initially collected material into context. We found various reasons for researchers to repeat the data collection step: getting a representative sample (I4: “so that [first analysis] was followed by collecting new data, using the methods identified in the literature and guided by insights from the earlier analysis”), getting context for a certain view (I5: “I needed to know the ethnic background of people [mentioned in tv-programs] for example”), and getting data to follow a new insight or direction (I3: “it was when I started having interviews that I realized that it is not so black and white and that is when the direction of my research changed completely”). These comments illustrate how analysis of the data leads to collection of additional data (I4, I5) and even completely new data (I3). These iterative steps suggest that initially researchers do not have specific criteria for data selection, only during analysis patterns emerge that lead to more specific search criteria. These changes in the search criteria manifest itself as a refinement in the research question.

After the new data has been collected another analysis follows, where, again, the material is inspected and relations between material are identified. The research question is further refined and will either lead to the next phase or another cycle of data collection and analysis. One researcher notes I1: “by organizing [the material] you create the story, I chose to use a chronological ordering, if I had organized my archive differently I would have written a different story. I could have organized it in supporters and opponents.” In this case a certain view on the data was chosen and the data organized accordingly. The effort it takes to reorganize the data for a certain view constrains the number of alternatives that are investigated, suggesting that having an overview of the alternatives early in the research cycle can be beneficial. Another researcher states I2: “and then you return again to the data and sometimes the literature, while part of the data has already been collected, because you feel that something is there but it does not come out.” In this case a certain contrast was expected in the data but did not appear after organization and analysis. This is another motivation for a more thorough exploration of views on the topic before analysis.

Presentation. The interviewees agree that at some point the data is fixed and the next phase starts. At this point a relevant sample of the data has been collected and this data is interpreted in the context of the refined research questions. This phase consists of analyzing and writing as the media studies researcher builds up a case to support his/her research questions by organizing the data and selecting appropriate qualitative evidence. After the analyses have been completed and the results have been written down, the findings are disseminated in publications and at conferences.

2.3 Relation to Other Models

We compare our model, as visualized in Figure 1, with previous work on describing the humanities research cycle. The model of literary critics describes six stages: preparation, elaboration, analysis and writing, dissemination, and further writing and dissemination [8]. The preparation stage corresponds with our exploration phase. The elaboration, and the analysis and writing stage correspond to our contextualization phase. While the dissemination, and further dissemination and writing stage corresponds to our presentation phase. That model, however, presents a cycle where all stages are completed in sequence while we find a number of shorter cycles in the initial stages. The shorter cycles serve to find a focus in the large amounts of available data. Another model proposes three phases: reading and annotation, developing interpretation, and presentation of interpretation [4], similar to our phases. The focus of this work, however, is on note taking and organizing processes and not on information seeking behavior. Additionally,

we provide more detail about the search processes in which media studies researchers engage. There is work on describing the scholarly practices of humanities researchers [27], but while this work describes the possible processes, e.g., organizing, it does not describe how these processes influence the research questions during the research cycle. The value of our model is that it makes the sequences of search processes and the gradual refinement of the research questions in the media studies research cycle explicit. Information needs of media researchers have been classified in four phases: getting an overview, identifying borderline exemplars, selection of specific programs and verification of facts [22]. These changes in information needs are consistent with our changes in the research question, but are not considered in the context of a research cycle and the possible repetition of information needs.

The relation of our model to search models [2, 10, 18], is that the stages and features associated with these models are observed in the media studies research cycle, but correspond to single processes, e.g., a data collection process. Our model describes the research cycle, the context of these search processes. The information behavior model by Wilson [38] provides a general framework to model users’ information behavior in context. We do not consider all possible variables of the model as it is too broad. Instead, we focus on the search processes (information seeking behavior) and the change of the information need in the research cycle due to learning about characteristics of the data (intervening variables).

2.4 Discussion

We find that the research cycle of media studies researchers is an iterative process, where processes of data collection and analysis alternate, see Figure 1. In the *exploration* phase an initial data set is collected and analyzed. The analysis of data is an expensive process that consists of extracting and organization of material, but is necessary to get insight into the material and to refine the research question. In the *contextualization* phase a new data collection process starts and another analysis follows. If this analysis reveals that certain patterns or views on a topic have been missed the process repeats. Otherwise a process of analysis, writing, and dissemination follows, i.e., the *presentation* phase. Factors causing media studies researchers to adjust their research question are discovering alternative views on a topic, and discovering trends in the data. These observations suggest that an interface that supports these types of discovery in an earlier stage of the research cycle will reduce the number of iterations necessary to arrive at a final research question and so support media studies research.

3. A SUBJUNCTIVE INTERFACE

The model of the research cycle in §2 provides insight in the requirements for a successful search interface for media studies researchers. Below we describe the development process of our subjunctive exploratory search interface, simply referred to as *subjunctive interface* in the remainder of the paper, followed by a detailed description of the interface features.

3.1 Development Procedure

In developing the subjunctive interface care has been taken to follow user centered design principles [5]. Here, we motivate the initial design of the subjunctive interface, describe the data used in our prototype, and findings from two rounds of usability testing.

Initial design. We established two requirements for an interface for media studies researchers: (i) to provide users with support for exploration, i.e., support in formulating queries, query refinement and exploring various aspects of a topic; and (ii) to provide support for discovering patterns in the data, i.e. to compare alternatives and

to observe trends in the data. A large body of work exists on interfaces for supporting exploratory search [6, 7, 20, 28, 37, 39]. Such interfaces provide support for the first requirement through visualizations, filters and facets. We start our interface development with the design of a prototypical exploratory search interface [6].

Not as well supported by this type of interface is the ability to compare alternatives. Subjunctive interfaces have been suggested for this purpose as this type of interface allows a user to perform multiple actions in parallel and compare the results, i.e., editing a document or searching a database. Typically multiple versions of a standard interface, e.g. a standard document editor, are presented side-by-side to create a subjunctive interface [23]. In web search, a multi-view interface has been proposed that supports multiple views of a traditional web search interface and allows users to explore more aspects of a topic than a single view variant [34]. Another requirement not as well supported is the discovery of trends in the data. Visualization laboratories, as those used in e-Science disciplines, are better suited for this purpose as these offer various visualization techniques, i.e., curves, scatter plots and renderings, to analyse large numerical datasets [14].

Given our requirements, we adapted the standard exploratory search interface design in two ways: (i) we extended the design to a subjunctive exploratory search interface by incorporating two side-by-side versions of a standard exploratory search interface; and (ii) we added a timeline visualization and a term statistics visualization in which the characteristics of the result sets obtained with each side of the interface are shown and can be compared.

Data set. Television studies (a sub-discipline of media studies) concerns the study of production, interpretation and/or reception of television. From an audiovisual archive we obtained a catalogue of about 1.5M television program descriptions to serve as data set to be accessed through our interface. We use descriptions as the actual programs are often not directly accessible due to copyright legislation [22]. The program descriptions are created by archivists, and primarily consist of metadata fields describing the program. For example, keywords, summary, and fields with program production information, e.g., broadcast date and program creator. The back-end of the interface consists of a Lucene SOLR index, where stopwords have been removed and stemming has been applied. For retrieval the Lucene implementation of the Vector Space Model is used.¹

Usability testing. In a first round of usability testing we presented a prototype of the subjunctive interface to two groups of media studies researchers, consisting of 12 and 16 subjects. A presentation with a walk-through of the interface was followed by a group interview. The three main findings are: (i) the importance of production information such as program broadcast date and program maker, next to the content of programs; (ii) program genre information is essential and a heavily studied subject in media studies; and (iii) the reception of television is often studied over time. We also received feature requests, i.e. the ability to exclude certain terms, to view the query history, and to load alternative archives such as news archives and television magazine collections.

After a new round of development we performed a usability study of the subjunctive interface. The subjects consisted of 30 first year information science students that participated as part of a class project. The main concerns of the subjects were with the cosmetics of the interface, the response time, and the size of the result snippets. After incorporating this feedback we performed a series of small pilot studies with media studies researchers to test the final design and to remove any further usability issues.

¹<http://lucene.apache.org/solr/>

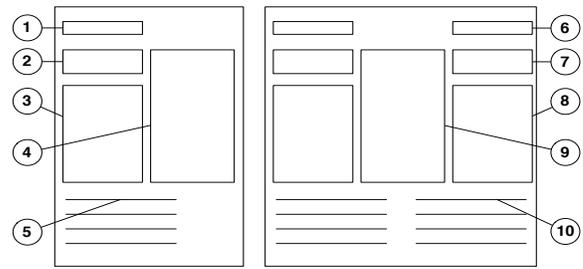


Figure 2: Schematic view of the baseline interface (left) and the subjunctive interface (right). Numbers are used for reference.

3.2 Subjunctive Interface Description

We start by describing an exploratory search interface that will serve as the basis for the subjunctive interface. In the remainder of the paper we refer to this interface as the *baseline* as it is used for comparison in our evaluation of the subjunctive interface described in §4. Figure 2 shows a schematic view of two interfaces: on the left-hand side the baseline and on the right-hand side the subjunctive interface. We use the numbers (1, ..., 10) in Figure 2 for reference specific components in the interfaces. [The interface is available at: ilps.science.uva.nl/MeRDES/usecase-ii.html](http://ilps.science.uva.nl/MeRDES/usecase-ii.html)

Baseline. The baseline interface consists of a search box (1), two filters: a timeline (2) and a term-cloud filter (3), a timeline and term statistics chart (4), and a result list (5). The baseline interface provides traditional search functionality in that typing a query in the search box (1) results in a ranked list of document snippets (5). Each result snippet describes a program with a title, broadcast date, and a maximum of 40 words from the summary of the program. Next to each snippet, a bookmark button is available. Bookmarking a program adds it to the query history, available as drop-down list, showing for each query the programs bookmarked in its result set. When clicking on a snippet an *overlay* with the complete program description appears. In the result set 25 program snippets are shown per page and the result set is limited to a maximum of 500 programs to keep the interface responsive.

The filters (2, 3) enable a user to rapidly refine the result set returned for a query [35]. Each subsequent filter that is applied operates on the remaining program descriptions. Filters are reset by issuing a new query or pressing a “clear filter” button. The timeline filter removes programs with a broadcast date outside of the selected range, see top Figure 3.

The term-cloud filter enables a type of faceted search over the



Figure 3: Screenshot of the timeline filter (top) and the term-cloud filter with the *people* facet selected (bottom).

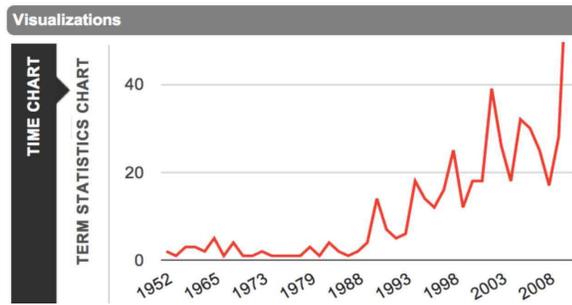


Figure 4: Timeline chart: y-axis showing the number of programs broadcasted per year; x-axis showing the years.

result set, see bottom Figure 3. Next to query refinement, faceted search also provides support for gaining insight in a topic [19, 35]. We decided on five facets based on the focus group interviews: *people* mentioned in a program, *makers* of a program, *channel* a program is broadcasted on, *words* are keywords characterizing the program, and *genre* of a program. Each cloud provides two modes of filtering: *retain* and *remove*. To *retain* a user clicks a term and only program descriptions that contain the term are kept. To *remove*, a user clicks a term and holds the mouse button, drags it slightly and releases the mouse button causing program descriptions that contain the term to be removed. Repeating an action deactivates a filter and “un-hides” documents affected by this filter.

The final parts of the baseline interface are the timeline and term statistics charts (4). These visualizations offer support for discovering trends in the data. They are not shown simultaneously, but are accessed through a *slide deck*. The timeline chart, see Figure 4, is shown by default. By clicking on the *term statistics slide* an animation shows the term statistics chart “sliding” over and covering the timeline chart. The timeline chart is subsequently accessible through the *timeline slide*. An example of the term statistics chart is shown in Figure 5. A drop-down menu allows the user to select one of the facets (people, maker, channel, words, and genre) to inspect the terms that occur most frequently in the program descriptions.

Subjunctive exploratory search interface. The left-hand side of the subjunctive interface consists of the same features as the baseline interface. On the right side the subjunctive interface further consists of an additional search box (6), timeline and term-cloud filter (7, 8), and result list (10); see Figure 2. The two search boxes with their respective filters and result lists are independent and in

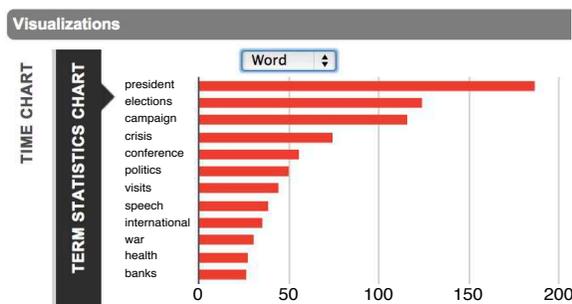


Figure 5: Term statistics chart: y-axis showing the terms with the highest frequency in the program descriptions of the current result set, x-axis showing the number of program descriptions that contain the term.

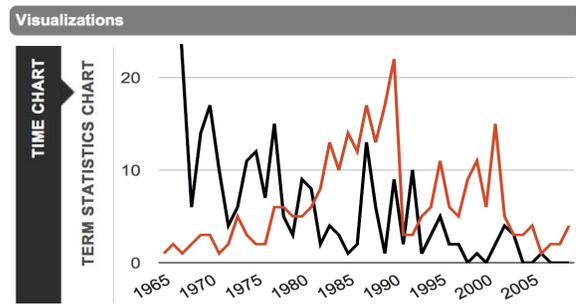


Figure 6: Subjunctive interface timeline chart: black corresponds to the result set for the query “protests”, the lighter shade (red) to the query “riots”. The axes are defined in Fig. 4.

essence provide the user with a second exploratory search interface. The visualizations in the subjunctive interface (9) differ from those in the baseline (4). The timeline chart shows two curves, one for each result set, see Figure 6. The curves are color coded black and red. Similarly, the search boxes are colored black and red on the left and right-hand side of the interface, respectively, to indicate their correspondence to the user. Analogously, the term statistics chart shows two bars per term, one for the frequency of the term in the left result set and one for its frequency in the right result set. The terms are required to occur in both result sets and are ordered by the total frequency in both sets, see Figure 7.

Note that although care has been taken in the design of the interface, we do not claim that this design is optimal. One suggestion for improvement is a tabbed view allowing a user to operate any number of instances of the interface and thus make any number of comparisons. Another issue arises from the term-cloud tabs, although offering faceted search the facets are hidden and only one facet is available at a time. Finally, we opted for cloud visualizations which have been found to be inferior to alphabetical listings [13]. The current design however satisfies the essential requirements for a subjunctive exploratory search interface and is suitable to answer our research questions.

4. USER STUDY

To assess the support provided by the subjunctive interface we conduct a user study with media studies researchers. Next, we describe the experimental design and our evaluation methodology.

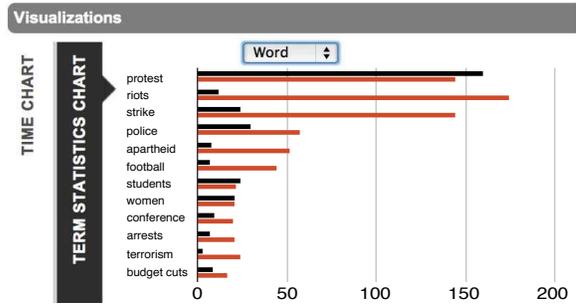


Figure 7: Subjunctive interface term statistics chart: y-axis showing terms with the combined highest frequency in the two result sets, x-axis showing the number of program descriptions that contain the term. Bar colors correspond to those in Fig. 6.

4.1 Experimental Design

The experiment was set up as a remote user study with a between subjects design [17]. We decided on a remote user study to be able to reach a wider audience of media studies researchers. A disadvantage is that there is less control over the setting of the study.

Study procedure. Subjects were recruited by spreading a URL among researchers and students at six media studies institutes. The URL directed subjects to a webpage explaining the experiment. Then subjects were presented with a consent form and a background questionnaire. Next, a three minute tutorial video of the interface was shown followed by a 3 minute practice session.

After practicing, subjects were given the following complex exploratory search task [16]: “As preparation for writing a research paper on the topic of migrants you investigate an audiovisual repository. You are interested in how migrants are represented on television. The goal of exploring the repository is to help you establish the initial research question for your paper.” Subjects were instructed to bookmark programs deemed relevant to formulate their research question and given 30 minutes to search for relevant programs with one of the two interfaces. After 30 minutes, or when subjects pressed a done button, a form was presented in which subjects were asked to submit a research question. While formulating their research question the subjects had access to the program descriptions bookmarked earlier during search. The final step consisted of a usability questionnaire. A session took about 45 minutes per subject. As a reward for participation, subjects received a 10 Euro gift certificate.

Subjects. The interface is developed to support media studies researchers, we therefore targeted subjects that had at least completed a Bachelor’s degree in media studies. Out of 61 subjects, 38 completed the experiment. Two subjects that did complete the experiment were excluded from the experiment: one subject that spent a total of 26 seconds interacting with the interface and one subject that had not yet completed a Bachelor’s degree. This left us with 36 subjects, 17 for the baseline and 19 for the subjunctive interface.

In terms of research experience, subjects are from a wide range of academic positions in media studies: 13 Master level students, 9 PhD students, 3 post doctoral researchers, 6 assistant professors, 1 full professor, and 4 research support staff. The research experience of subjects, in terms of the median (*MD*) and interquartile range (*IQR*), varies as subjects are a mix of researchers and students (*MD* = 3, *IQR* = 0-6.5). We asked subjects background questions using a 5 point Likert-type scale, where a one indicates no agreement and a five indicates extreme agreement. Subjects generally reported high levels of experience in general computer use (*MD* = 4, *IQR* = 4-5) and using online search tools (*MD* = 4, *IQR* = 4-5). Subjects had little previous experience with the topic of the search task, e.g., media and migration (*MD* = 2, *IQR* = 1-3). We found no significant differences between the groups in terms of these statistics.

4.2 Evaluation Methodology

We assess the support the subjunctive interface provides for media researchers in terms of three aspects: (i) support in exploration of different views of a topic; (ii) support in refining a research question; and (iii) general usability.

Exploration of different views. It is difficult to obtain a fixed set of relevance judgements for a complex exploratory search task, e.g., gathering documents that serve as basis to formulate a research question, as relevance is hard to determine in such a broad task [17]. Instead of using precision and recall, we evaluate support for exploration in terms of user interaction derived from server side log files. We hypothesize that an interface that provides better support for

Table 1: Medians and interquartile ranges for user interactions with of the baseline (bl) and subjunctive (sj) interface.

interface feature	bl	sj
query formulations	3 (2-6)	5 (3.3-7.8)
bookmarks	9 (5.8-22.3)	9 (2.8-12.8)
document views	3 (1-7)	2 (0.3-5.3)
timeline filter	3.5 (2.5-9.5)	3 (0-8.5)
term-cloud filter	28 (13.5-41)	16 (9-36)
filter/analysis time (sec)	303 (204.8-589.3)	384 (222.5-544.3)
inspect result time (sec)	253 (51.3-438.8)	202 (94.8-385.3)
total time (sec)	532 (308-995)	575 (386.3-947.3)

exploration will enable subjects to generate more query formulations and that subjects will bookmark more diverse documents.

Research question refinement. To evaluate subjects’ research questions we asked three media researchers, experts in the field of media and migration, to act as assessors: an associate professor (judge₁), a full professor (judge₂), and a post-doctoral researcher (judge₃). Research questions were judged on five criteria: (i) general quality (g); (ii) extend to which a scope is defined, i.e., limiting the question to a certain person or time (s); (iii) clarity of formulation (f); (iv) embedding (e), i.e., the degree in which the research question relates to literature; and (v) originality (o). In the media studies research cycle the refinement of the research question is a manifestation of changes in data selection criteria. We therefore hypothesize that research questions formulated by subjects with the subjunctive interface will be judged higher on the scope criterion.

Usability. Exploratory search systems are more complex than standard web search interfaces [35]. We introduce a subjunctive version of an exploratory search interface that essentially doubles the amount of features in the interface. We assess the subjunctive interface in terms of usability and use the following criteria: (i) usefulness, (ii) intuitiveness; (iii) ease of use; and (iv) interestingness, based on [21].

All judgements regarding the research questions and questions in the exit-questionnaire are given on a five point Likert-type scale, where the level of agreement is indicated in the range from one (not at all) to five (extremely). When reporting results the Wilcoxon ranksum test is used to determine significant differences between groups at the $\alpha < .05$ level. In tables significant differences are always in comparison to the baseline and indicated in bold face.

5. RESULTS

5.1 Exploratory Search Support

To answer our first research question we evaluate the performance of the baseline (bl) and subjunctive (sj) interface on a complex exploratory search task in terms of user interaction statistics and in terms of search patterns.

User interaction statistics. Our first hypothesis states that with the subjunctive interface subjects will formulate more queries and bookmark more diverse documents. We first compare subjects’ interactions with the baseline and subjunctive interface, followed by an analysis of the diversity of the bookmarked documents.

Table 1 shows the medians (*MD*) and interquartile ranges (*IQR*), for interactions of subjects with features of the two interfaces. We find that the number of query formulations is higher for subjects using the subjunctive interface (bl *MD* = 3, sj *MD* = 5). The difference is significant as indicated by a Wilcoxon ranksum test ($W = 255.5, p < .05$) indicating that the subjunctive interface provides

more support for generating new query formulations. We observe that subjects bookmark a similar number of documents with the interfaces (bl $MD = 9$, sj $MD = 9$). The high-end of the interquartile range for the baseline is higher but the difference is not significant. That subjects do not bookmark more documents may be due to the task description, i.e., the goal is to formulate a research question and not to bookmark as many relevant documents as possible.

The remaining interaction statistics demonstrate no apparent differences. We note that subjects spend a similar amount of time searching (bl $M = 532$, sj $M = 575$), but that users of the subjunctive interface spend more time operating the filters and/or analysing the visualizations (bl = 303, sj = 384). While in the baseline interface more time is spent inspecting results (bl = 254, sj = 202), i.e., reading result snippets and viewing documents. Although the difference is not significant it is to be expected that subjects spend more time analysing and filtering with the subjunctive interface as more information is presented. That subjects spent less time, or a similar amount of time inspecting results is surprising as the subjunctive interface presents twice as many results. We look further into this when analysing the interaction patterns.

Next we investigate whether there are differences in the diversity of bookmarked documents. We use cosine similarity as a distance measure and calculate the average pairwise cosine similarity of the documents bookmarked (D_s) by a subject (s):

$$avg_sim(s) = \frac{1}{|P_s|} \sum_{(d,d') \in P_s} sim(d,d'),$$

here P_s is the set of pairs of documents bookmarked by a subject: $P_s = \{(d,d') : d,d' \in D_s, d \neq d'\}$ and $sim(d,d')$ is defined as:

$$sim(d,d') = \frac{\sum_{i=1}^n d_i \cdot d'_i}{\sqrt{\sum_{i=1}^n (d_i)^2} \cdot \sqrt{\sum_{i=1}^n (d'_i)^2}}.$$

The average similarities for subjects using the baseline ($avg_sim MD = .62$, $IQR = .56-.69$) are higher than the similarities of subjects using the subjunctive interface ($avg_sim MD = .52$, $IQR = .43-.63$) and this difference is significant ($W = 195 p = .05$). That documents bookmarked with the subjunctive interface are less similar than those bookmarked with the baseline indicates that with the subjunctive interface a more diverse set of documents are explored.

User interaction patterns. The interaction statistics suggest that the subjunctive interface provides more support for exploration than the baseline, if so we would expect to find differences in subjects' search behavior. We first describe the process of creating the interaction patterns based on maximal repeating patterns [29] and then analyze the patterns generated with the two interfaces.

An interaction pattern consists of all of a subject's search actions during a search session. We identify the following action types: submitting queries (Q), using filters (F), inspecting results (I), bookmarking (B), viewing program descriptions (D), paginating to new result pages (P), and closing the interface (S). Repeated actions are aggregated into a single action type, e.g., queries submitted in the left or right search box of the subjunctive interface are considered as a single query (Q) action. The purpose of the resulting interaction pattern is to reveal transitions between interaction types. Per subject group all occurrences of possible sub-patterns of at least two subsequent actions are counted to find the maximal repeated patterns (MRP) for each interface. For example, the sequence of actions: QFIFIQFI, contains the following MRP: QFI and FI, as these are the longest sequences that are repeated.

Table 2 shows the top 8 MRPs that start with a query action (Q), a filter action (F), or a bookmark action (B). In both interfaces the most frequent transition after submitting a query is to inspect the results (bl $QI = 37$, sj $QI = 57$). After this initial behavior, subjects using the subjunctive interface more often reformulate their query

Table 2: Users' most frequent maximal repeated patterns with the baseline (bl) and subjunctive (sj) interface. Here Q is submitting queries, F is using filters, I is inspecting results, B is bookmarking, D is viewing documents, P is paginating to reach new result pages, and S is closing the interface.

Q starts pattern		F starts pattern		B starts pattern							
#	bl	#	sj	#	bl	#	sj				
9	QFIF	12	QFQ	12	FIQ	18	FIFIFIF	8	BPBP	6	BIQ
9	QIQI	12	QIFI	14	FIFIF	19	FQ	10	BPB	7	BS
12	QIQ	13	QIF	15	FQ	23	FIB	11	BDB	8	BIB
13	QIFI	18	QIQI	15	FID	29	FIFIFI	13	BD	11	BDB
15	QIF	22	QIQ	25	FIB	34	FFIF	14	BFI	11	BFI
27	QFI	23	QFI	30	FIFI	55	FIFI	15	BP	14	BF
32	QF	37	QF	37	FIF	62	FIF	15	BI	18	BD
37	QI	57	QI	97	FI	116	FI	16	BF	22	BI

(bl $QIQ = 12$, sj $QIQ = 22$), while users of the baseline prefer filtering (bl $QIF = 15$, sj $QIF = 13$). This is consistent with the earlier finding, see Table 1, that the subjunctive interface provides more support for formulating new queries.

When starting with a filtering action and then inspecting the results (FI) subjects using the subjunctive interface tend to transition more from filtering actions to bookmarking or viewing documents (bl $FIFI = 30$, sj $FIFI = 55$). With the subjunctive interface subjects spend more time refining and inspecting result snippets (sj $FIFIFI = 29$); the number of times a filter and inspection sequence leads to a bookmark is comparable (bl $FIB = 25$, sj $FIB = 23$). The extra time spent refining and inspecting can be explained by the presence of the second result set in the subjunctive interface as subjects have a larger set of program descriptions at their disposal.

In the baseline interface a bookmark action is often followed by moving to the next result page (bl $BP = 15$) and (bl $BPBP = 8$), while in the subjunctive interface more often program descriptions are viewed (sj $BD = 18$). This suggests that subjects using the baseline interface are unable to formulate new queries or use filters to refine the result set and resort to browsing more result pages in the result set. This is consistent with the observation in the interaction statistics, see Table 1, that with the baseline more time is spent inspecting results; this is similar to the behavior observed in web search when users face a difficult search task [1]. Users of the subjunctive interface on the other hand, tend to bookmark documents on the first result page (sj $BIB = 8$) and (sj $BDB = 11$).

We have determined that there are differences in the interaction patterns of subjects using the baseline and subjunctive interface. Interaction patterns show that with the subjunctive interface subjects alternate more between formulating queries and inspecting results than subjects using the baseline interface. We also find that users of the baseline are more often unable to refine their information need and resort to an exhaustive search of the results.

5.2 Research Question Formulation Support

To address our second research question we evaluate how the exploration provided by the two interfaces affects the research questions formulated by the subjects. We perform two types of evaluation: (i) a quantitative evaluation where we use explicit judgements of the research questions; and (ii) a qualitative analysis where we divide research questions into phrases and classify these into several types to compare the composition of the research questions.

Research question formulation performance. Our second hypothesis states that research questions formulated by subjects with the

Table 3: Medians and interquartile ranges for judgements of the research questions on five criteria: general quality (g), scope (s), formulation (f), embedding (e), and originality (o), by three media researchers for the baseline (bl) and subjunctive (sj) interface.

	judge ₁		judge ₂		judge ₃	
	bl	sj	bl	sj	bl	sj
g	3 (2-4)	4 (3-4)	4 (2-4)	3 (2.3-4)	4 (3.8-4)	4 (3.3-5)
s	3 (2-4)	4 (3-4)	2 (2-4)	2 (2-3)	4 (3-4)	4 (2.3-4)
f	3 (3-4)	4 (2.3-4)	3 (2-4)	3 (2.3-4)	4 (4-4.3)	4 (4-4.8)
e	3 (3-4)	3 (2.3-4)	2 (2-3)	3 (2-3.8)	3 (1.8-4)	3 (2-4)
o	4 (3-4)	4 (3-4)	4 (3-4)	4 (3-4)	3 (2-4)	4 (2-4)

subjunctive interface will be judged higher on the scope criterion. The top of Table 3 shows the medians and interquartile ranges for the judgements of the research questions on the five criteria, i.e., general quality (g), scope (s), formulation (f), embedding (e), and originality (o), by three media researchers. Overall agreement of the assessors on the criteria is low as indicated by Fleiss’ Kappa ($\kappa < .2$). Agreement is stronger on the scope criterion ($\kappa = .2362$).

Subjects’ research questions are judged to be good in terms of quality, formulation, and originality, for all three assessors ($MD \geq 3$). The level of embedding is lower ($MD \leq 3$), as researchers were unable to consult any literature during the experiment. The judgements for the scope of the research questions, are mixed. We observe, however, no apparent differences between the baseline and subjunctive interface for the judgement criteria.

That we observe no differences may be due to the open ended nature of the task of formulating a research question and the difficulty of judging research questions without any further context.

Research question composition. The research questions formulated by the media studies researchers provide a rich source of qualitative data. To investigate if there is a difference in the views and topics of the research questions, we perform a qualitative analysis of the phrases that define the scope in the research questions of the 36 subjects. We identify five types of phrases: defining (i) a program genre; (ii) a group to study; (iii) a focus on a theme or person; (iv) a time period; or (v) a part of the program production process.

Table 4 shows the number of occurrences of types of phrases that determine the scope of a research question. A similar number of phrases of type genre, period, and production is found in research questions generated with the baseline and subjunctive interface. There is a difference in the number of phrases of type group (bl = 14, sj = 17), but not in terms of unique phrases. Here, the phrase “migrants” is used to specify the population of study, most likely influenced by the use of the term in the search task. We observe that in research questions generated with the subjunctive interface more often phrases that describe a specific theme occur as compared to the baseline (bl = 8, sj = 12) and in most cases these themes are unique (bl = 7, sj = 11). This suggests that the subjunctive interface provides subjects with more support to explore different themes surrounding a topic and that they use this information to scope their research question. There is little influence on other types that determine the scope of a research question, e.g., a time period or television genre, as both interfaces provide users with the ability to spot trends through the timeline and term statistics chart.

Next we provide an example-based comparison to further illustrate the effect of the interfaces on the scope of the research questions. Table 5 shows the top 3 research questions for the baseline and subjunctive interface at the top and bottom respectively. Questions are

Table 4: Number of occurrences (unique occurrences) of types of phrases that determine the scope of a research question.

interface	genre	group	theme	period	production
baseline	11 (6)	14 (8)	8 (7)	10 (10)	2 (2)
subjunctive	10 (7)	17 (8)	12 (11)	9 (9)	3 (3)

ranked in terms of the sum of the three assessors’ judgements on the scope criterion. We observe that when subjects use the baseline, questions are based on observations of trends in the timeline, i.e., the increase and decrease in the occurrence of a query term in programs. This leads to research questions that focus on a single topic during a certain period, i.e., the representation of the Islam (bl rq₁) and the representation of a political figure (bl rq₂). The ability to compare changes in frequency of query terms on a timeline in the subjunctive interface, however, inspires subjects to consider more views. This leads to research questions that include interactions between multiple aspects of a topic, e.g., representation of Muslims and the influence of terrorism (sj rq₁) and difference in representation of refugees’ children in fictional programs compared to documentaries. In the last research questions (bl and sj rq₃) the effect of the subjunctive interface is most obvious. Although both questions follow the change in use of terminology over time, in the subjunctive interface a contrast is made between two terms.

5.3 Usability

To answer our third research question we look at the usability of the interfaces. Table 6 shows the medians and interquartile ranges of subjects’ judgements of the usability of the two interfaces. Subjects indicate that both are intuitive (bl = 4, sj = 4) and that they are not difficult to use (bl = 2, sj = 2). Of the subjects, 75% do not find the subjunctive interface difficult to use (difficult ≤ 3). This suggests that subjunctivity can be added to exploratory interfaces with little cost to the difficulty and intuitiveness of the system. We further asked subjects if the interfaces were interesting to use and useful for media research. Subjects indicate that both interfaces are interesting (bl = 4, sj = 4) and useful, in case of the subjunctive interface subjects indicate it to be extremely useful (bl = 4, sj = 5). Regarding the visualizations subjects indicate a preference for the timeline chart. We suspect that the information in the term statistics chart is more difficult to interpret and therefore used less.

6. DISCUSSION

Regarding our first research question, we find significant evidence that with the subjunctive interface, media studies researchers exhibit different search behavior than with the baseline interface on a complex exploratory search task. Subjects are able to formulate more queries and bookmark more diverse documents than with a traditional exploratory search interface. Inspection of the interaction patterns confirms these findings. Users of the subjunctive interface follow a pattern of reformulating a query and inspecting results followed by another query reformulation and result inspection.

Table 6: Medians and interquartile ranges for the usability of the baseline (bl) and subjunctive (sj) interface.

question	bl	sj	question	bl	sj
intuitive	4 (4-4)	4 (3-4)	difficult	2 (2-2)	2 (2-3.8)
interesting	4 (4-5)	4 (4-5)	useful	4 (4-5)	5 (4-5)

Table 5: Top 3 research questions for the baseline (left) and subjunctive interface (right), ranked in terms of the scope criterion. Alterations to the research questions are indicated by [...] and serve to protect the anonymity of subjects or to improve clarity.

rq ₁ :	How is the Islam represented in factual television genres during the period 2000-2010?	How are Muslim immigrants represented on television and what is the role of terrorism in this representation. Case study of several episodes of [program _A] and [program _B] about Muslims in [country].
rq ₂ :	How is [political figure] represented by the public broadcasting corporation during the elections for the house of representatives in 2010?	How are the experiences of refugees' children represented on television in fiction and documentaries from 1990 to 2010? Do we find any differences or changes and can these be explained?
rq ₃ :	Investigation of the evolution of the term integration in news and human interest programs broadcasted between 1992 and 2012.	In 1987 we observe a diminishing in the use of the term migrant worker and a rise in the use of the term immigrant. Is it possible to identify a cause in the broadcasting schedule of that time? Are there specific programs that started this development?

tion, while users of the traditional exploratory search interface formulate less queries and look through more result pages.

With regard to our second research question, we find that with both interfaces researcher are able to formulate high quality research questions. A qualitative analysis of the research questions shows that there is a difference in the research questions that subjects formulate. With the subjunctive interface subjects use more diverse themes to scope their research question. There is no influence on other types of defining the scope, e.g., a time period, as both interfaces provide users with the ability to spot trends in visualizations. An example based comparison of the top 3 research questions in terms of scope illustrates the difference in the number of views on a topic incorporated in the research questions.

Turning to the third research question, we find that although the complexity of the subjunctive interface in terms of features almost doubled compared to the standard exploratory search interface, most users indicate that the subjunctive interface is intuitive and not difficult to use. Users indicate that the subjunctive interface is interesting and judge it to be extremely useful for media research.

The subjunctive interface was developed to support the exploration phase in the media studies research cycle by supporting multiple views on a topic and discovering trends in the data. The above findings demonstrate that a subjunctive exploratory search interface can indeed provide this type of support for media studies research. A limitation of this study is that the time and data restrictions in the experiment make it an abstraction from the real research cycle; a longitudinal study where the subjunctive interface is used by media studies researchers will have to be conducted. Another limitation is that the study of the research cycle is based on a small number of interviews. The development of the subjunctive interface, however, has been done in close co-operation with media studies researchers. Therefore the model and the ideas on which the subjunctive interface has been based, have repeatedly been under close examination from the targeted end users, thereby strengthening its foundation.

7. RELATED WORK

The work in this paper touches on various areas of research and we briefly highlight related work. Many specific related publications have already been cited earlier in the paper. We complement this with a brief and slightly more general discussion. As to exploratory search [24], many exploratory search tools have been suggested [6, 7, 11, 12, 20, 28, 37, 39] which served as inspiration in the development of our interface.

Tools in the humanities exist for the analysis of data once it has been collected. Some of the more recent projects are: MONK² a data analysis laboratory that provides visualizations and data mining tools; TaPoR³ an indexing and search tool to locate co-occurring phrases in texts; and Pliney [4] a tool for note taking and organizing

²<http://monkproject.org/>

³portal.tapor.ca/

to support insight generation. The Google ngram viewer is a tool that enables comparison of trends in term occurrences on a timeline in a large book collection supporting cultural analytics [26]. A limitation of these tools is the lack of support for exploration of different subsets of documents, e.g., query result sets.

Subjunctive interfaces have been proposed in a more general setting [23], e.g., to compare two documents in a side-by-side editor. For web search an interface with multiple views of a traditional web search interface was shown to be effective to support complex search tasks [34]. We extend this body of work by investigating the value of multiple views in a exploratory search interface.

8. CONCLUSION

In this paper we have presented a subjunctive exploratory search interface to support media studies research. By analysing the media studies research cycle we have found that media studies researchers require support in discovering multiple views on a topic and discovering trends in data to refine their research question in an earlier stage of the research cycle. We have developed a subjunctive exploratory search interface and performed a user study to assess its value for media studies researchers. We have found that with the subjunctive interface media studies researchers are able to formulate more queries and bookmark more diverse documents compared to a standard exploratory search interface. In a qualitative analysis of the research questions formulated by media studies researchers we have found evidence to suggest that the influence of the subjunctive interface is predominantly on the scope of the research question. Specifically, users of the subjunctive interface incorporate more views on a topic in their research question than users of the standard exploratory search interface. We have observed no advantage for other types of defining the scope as visualizations in both interfaces enable spotting trends in the data. In terms of usability, media studies researchers report that the subjunctive interface is intuitive and not difficult to use, suggesting that the additional complexity in terms of features in the subjunctive interface does not reduce its usability. These findings are important for media studies, but also have further implications. The continuous increase in data and its effect on the importance of exploration in the research cycle indicates a need to study this effect in other areas of the humanities. Additionally, with respect to exploratory search, the support provided by the subjunctive interface for exploration suggests a new direction for the development of exploratory search systems. In future work we plan to launch the subjunctive interface at a television archive to assess its value for media studies in a longitudinal study.

9. ACKNOWLEDGMENTS

This research was supported by the European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme, CIP ICT-PSP under grant agreement nr 250430 (GALATEAS), by the PROMISE Network of Ex-

cellence co-funded by the 7th Framework Programme of the European Commission, grant agreement no. 258191, by the DuOMAN project carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments under project nr STE-09-12, by the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.066.512, 612.061.814, 612.061.815, 640.-004.802, 380-70-011 and by the Center for Creation, Content and Technology (CCCT).

References

- [1] A. Aula, R. Khan, and Z. Guan. How does search behavior change as search becomes more difficult? In *SIGCHI'10*, pages 35–44, 2010.
- [2] M. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5): 407–424, 1989.
- [3] C. Borgman. The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4), 2009.
- [4] J. Bradley. Thinking about interpretation: Pliny and scholarship in the humanities. *Literary and linguistic computing*, 23(3):263–279, 2008.
- [5] T. Brinck, D. Gergle, and S. Wood. *Usability for the Web: designing Web sites that work*. Morgan Kaufmann, 2002.
- [6] R. Capra and G. Marchionini. The relation browser tool for faceted exploratory search. In *JCDL'08*, page 420, 2008.
- [7] R. Capra, G. Golovchinsky, B. Kules, D. Russell, C. Smith, D. Tunkelang, and R. White. HCIR 2011. In *SIGIR Forum*, volume 45, pages 102–107, 2012.
- [8] C. Chu. Literary critics at work and their information needs: A research-phases model. *Library & Information Science Research*, 21(2):247–273, 1999.
- [9] E. Collins and J. Michael. How do researchers in the humanities use information resources? *Liber Quarterly*, 21(2), 2012.
- [10] D. Ellis and M. Haugan. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *J. Doc.*, 53(4):384–403, 1997.
- [11] G. Golovchinsky and J. Pickens. Interactive information seeking via selective application of contextual knowledge. In *IiX'10*, pages 145–154, 2010.
- [12] M. Hearst. Tilebars: visualization of term distribution information in full text information access. In *SIGCHI'95*, pages 59–66, 1995.
- [13] M. A. Hearst and D. Rosner. Tag clouds: Data analysis tool or social signaller? In *HICSS '08*, pages 160–169, 2008.
- [14] T. Hey and A. Trefethen. The data deluge: An e-science perspective. In *Grid computing*, pages 809–824, 2003.
- [15] P. Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *J. Doc.*, 52(1): 3–50, 1996.
- [16] K. Järvelin and P. Ingwersen. Information seeking research needs extension towards tasks and technology. *Information Research*, 10(1):10–1, 2004.
- [17] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *FTNIR*, 3(1-2):1–224, 2009.
- [18] C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *JASIS*, 42(5):361–371, 1991.
- [19] B. Kules, R. Capra, M. Banta, and T. Sierra. What do exploratory searchers look at in a faceted search interface? In *JCDL'09*, pages 313–322, 2009.
- [20] Y. Lin, J. Ahn, P. Brusilovsky, D. He, and W. Real. Image-sieve: Exploratory search of museum archives with named entity-based faceted browsing. *ASIST'10*, 47(1):1–10, 2010.
- [21] A. Lund. Measuring usability with the use questionnaire. *Usability and User Experience*, 8(2):8, 2001.
- [22] B. Lunn. User needs in television archive access: Acquiring knowledge necessary for system design. *JoDI*, 10(6), 2009.
- [23] A. Lunzer and K. Hornbæk. Subjunctive interfaces: Extending applications to support parallel setup, viewing and control of alternative scenarios. *ACM TOCHI*, 14(4):17, 2008.
- [24] G. Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, 2006.
- [25] L. Meho and H. Tibbo. Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *JASIST*, 54(6):570–587, 2003.
- [26] J. Michel, Y. Shen, A. Aiden, A. Veres, M. Gray, J. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, et al. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176, 2011.
- [27] C. Palmer. Scholarly work and the shaping of digital access. *JASIST*, 56(11):1140–1153, 2005.
- [28] A. Shiri. Metadata-enhanced visual interfaces to digital libraries. *JIS*, 34(6):763–775, 2008.
- [29] A. Siochi and R. Ehrich. Computer analysis of user interfaces based on repetition in transcripts of user sessions. *ACM TOIS*, 9(4):309–335, 1991.
- [30] A. Strauss and J. Corbin. *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc, 1990.
- [31] E. Toms and H. O'Brien. Understanding the information and communication technology needs of the e-humanist. *J. Doc.*, 64(1):102–130, 2008.
- [32] J. Unsworth. Scholarly primitives: what methods do humanities researchers have in common, and how might our tools reflect this. *Symposium on Humanities Computing: Formal Methods, Experimental Practice*, 2000.
- [33] J. Unsworth. Tool-time, or 'haven't we been here already?' ten years in humanities computing. *Transforming Disciplines: The Humanities and Computer Science*, 2003.
- [34] R. Villa, I. Cantador, H. Joho, and J. Jose. An aspectual interface for supporting complex search tasks. In *SIGIR*, pages 379–386, 2009.
- [35] R. White and R. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.
- [36] M. Wilson, R. White, et al. Evaluating advanced search interfaces using established information-seeking models. *JASIST*, 60(7):1407–1422, 2009.
- [37] M. Wilson, B. Kules, et al. *From Keyword Search to Exploration*. Now Publishers, 2010.
- [38] T. Wilson. Models in information behaviour research. *J. Doc.*, 55(3):249–270, 1999.
- [39] K. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *SIGCHI'03*, pages 401–408, 2003.