Exploratory Search Pipes with Scoped Facets

Tim Gollub
Bauhaus-Universität Weimar
Weimar, Germany
tim.gollub@uni-weimar.de

Leon Hutans
Bauhaus-Universität Weimar
Weimar, Germany
leon.hutans@uni-weimar.de

Tanveer Al Jami
Bauhaus-Universität Weimar
Weimar, Germany
tanveer.al.jami@uni-weimar.de

Benno Stein
Bauhaus-Universität Weimar
Weimar, Germany
benno.stein@uni-weimar.de

ABSTRACT

This paper presents faceted search technology tailored to the peculiarities of exploratory search tasks. In contrast to traditional faceted search systems, facets in our system are arranged as a pipe, i.e., are applied sequentially one after the other. Moreover, the facets in a pipe are not applied throughout the whole sequence, but are limited to a user-definable scope, such that the search query can be broadened by adding a facet to the pipe. By this modification, the user interaction with our exploratory search engine resembles rather an open walk through the facet space spanned by the connections between documents and facets, than a progressive filtering of the document collection. We argue that this shift in the interaction paradigm is a much better fit to exploratory search scenarios. A user study with a prototype implementation attests an improved usability of our system compared to traditional faceted search systems for complex exploratory search tasks.

CCS CONCEPTS

• **Information systems → Information retrieval**.

KEYWORDS

exploratory search; faceted search; scoped facets

ACM Reference Format:
Tim Gollub, Leon Hutans, Tanveer Al Jami, and Benno Stein. 2019. Exploratory Search Pipes with Scoped Facets. In *2019 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '19), October 2–5, 2019, Santa Clara, CA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341981.3344247

1 INTRODUCTION

This paper introduces scoped facets as a concept to improve the exploratory search support of faceted retrieval systems. We consider the search task of a user to be exploratory, if the user's goal is to learn about the contents of a document (sub-) collection, rather than finding a specific piece of information. In other words, exploratory searchers ask for the information needs that could be answered on the basis of a specific document collection more than they ask for the documents that could answer a specific information need.

For example, a sub class of exploratory search is serendipity search [14], where the users' goal is to find interesting topics to search for in subsequent focused search tasks. More elaborated exploratory search tasks arise in the context of digital humanities research, computational social science, or investigative journalism, where corpora are explored with the goal of discovering relevant research questions (cf. distant reading [11]). Even further, exploratory search tasks arise from vague memories [2], where users have to retrieve forgotten keywords which are necessary to formulate a focused search query, or because users want to attest the lack of relevant documents in a collection, for which all reasonable search queries have to be found and tested (cf. negative search [6]).

Besides query suggestions, which allow users to get a glimpse into the information needs supported by a search engine, the use of facets as a paradigm to facilitate learning about and browsing through the contents of a document collection has been proposed to support exploratory search tasks (see Section 2). By presenting a set of relevant facets next to the search results for a query, so the intuition, exploratory searchers can start their search with a broad query, and then proceed with selecting facet terms to progressively narrow down into smaller result sets. While this search scenario covers an essential activity of exploratory searchers, we like to argue in line with White and Roth [16] that supporting the opposite scenario, starting out from a small result set, then broadening the query to refocus on related but different document subsets, is at least equally important.

To improve the exploratory search experience of users in this regard, we introduce the concept of scoped facets in conjunction with exploratory search pipes in Chapter 3. In contrast to traditional faceted search systems known e.g. from e-commerce sites or digital libraries, our proposal is to arrange facets selected by the user sequentially in a *pipe*. At each position of the pipe, the user can choose which prior facets should be evaluated to determine the relevant documents at this position. If this *facet scope* does not comprise all prior facets, the resulting search query is broadened, potentially leading to documents not present in previous exploration steps. With the concept of facet scopes, the user can conveniently switch between narrowing down and broadening the current search query. And since even broadening the query means progressing in the

ICTIR '19, October 2–5, 2019, Santa Clara, CA, USA

ACM ISBN 978-1-4503-6881-0/19/10…$15.00
https://doi.org/10.1145/3341981.3344247

pipe, the full exploration path is retained over the whole search session, allowing the user to reflect on the search directions taken.

The usability of our proposal is evaluated by the means of a user study, where we ask participants to solve exploratory search tasks with a prototypical implementation of our concept and a classical faceted search system as baseline. The design and results of this study are presented in detail in Section 4. While for simple search tasks the overall usability of both systems is assessed equally as being of average quality, the usability differs significantly for complex exploratory search tasks. Here, the usability score for our system increases from average to good, while the usability score for the baseline system decreases from average to bad.

2 RELATED WORK

Exploratory search had been coined as an information retrieval topic around 2005, where a series of workshops were organized that eventually led to the seminal book publication "Exploratory Search: Beyond the query-response paradigm" by White and Roth in 2009 [16]. From a historical perspective, exploratory search picked up on information retrieval problems such as subject search [1] and search result clustering [5], which address broad or vague information needs of users by providing topical overviews over retrieved document sets. Exploratory search departures from these problems in the respect that no longer the focus is on supporting users finding relevant documents, but is generalized to the task of supporting users learning about and investigating the contents of a specific document collection (including "which subtopics are addressed in my search results?" as one example) [10].

Over the years, a variety of exploratory search systems, like the Flamenco browser [18], the mSpace Explorer [13], the Relation Browser [3], Querium [7], or SearchLens [4] have been presented. A commonality of all these exploratory search systems is that they employ facet-like structures as their primary exploratory search feature. Concerning our proposal of exploratory search pipes with scoped facets, the mSpace Explorer represents the most relevant prior work, as it also implements the idea of a sequential arrangements of facets. However, without the concept of scoped facets, adding facets to the sequence may only result in more specific queries, hence the added value over traditional faceted search systems concerning exploratory search support is limited in our view.

A further stream of relevant research originates from the semantic web community, where faceted search systems have meanwhile been proposed for the exploration of ontologies (see [15] for a recent survey). The systems most related to our work from this stream of research are gFacet [8], and the proprietary semspect system[1]. In these two systems, the facets selected by a user are laid out not as a sequence but as a graph, visualized on a two-dimensional plane. Though the representation of the exploratory search path as a graph is more powerful in the sense that it allows branching at any position of the path, the two-dimensional layout makes displaying and interacting with (especially long) facet terms challenging, which is why we opted for a linear sequence, instead. More importantly, as holds for the mSpace Explorer, also these two systems do not consider the broadening of a search query by reducing the facet scope, but always apply all facets that share a common path.

[1]http://semspect.de/

3 APPROACH

In this section, we formalize the concept of exploratory search pipes with scoped facets, and present how this concept is implemented in the prototype system *podascope* that we use in our evaluation. We roughly follow the notation of Sacco and Tzitzikas [12], and define a faceted search system as the quadruple $S = ((T, \leq), D, I, Q)$, where $(T, \leq)$ is a taxonomy which itself consists of a terminology T with a root term, as well as a partial order $\leq$ over T called subsumptions, which defines the parent-child relationships between the terms in T. Every term in T is represented by a term id, a label, as well as an optional description. Terms that are subsumed only by the root node are called facets, and belong to the facet set $F \subset T$. Further on, D denotes the document collection that can be explored by our system. Each document in D consists of fields, i.e., key-values pairs with keys such as "authors", "title", "entities", etc.. For all keys which refer to a term in T, we take all unique values and add them as leaf terms to the taxonomy. From these values, we also infer the relations between the documents in D and the terms in T. These term-document relations are denoted by I, a function $I : T \rightarrow 2^D$ commonly referred to as interpretation. Ultimately, Q is the set of all queries that can be formulated over T using standard boolean operators.

Faceted Search. The starting point for our formalization is a user who wants to explore the subset of documents D_q that are relevant for a user specified search query q. To allow the user to issue keyword based search queries, our prototype initially provides a search box for keyword based searches within the fields of D. To facilitate faceted search, the subset of facets F which are associated with D_q via $\leq$ and I are displayed next to the search results. A screenshot of our prototype, where the remaining facets available for a search query are shown in the top most element, is provided in Figure 1. Users can browse through the facets by clicking on the navigation element next to the child count of a facet. The current browsing path is displayed left hand to a term search box, which can also be used to move back up in the taxonomy. By selecting a facet term, the user can, as usual, add the term to the current search query.

Exploratory Search Pipes. With the concept described so far, the user can search and browse through the relevant facets, but cannot select terms from different facets and explore the relations between them. To this end, we introduce the concept of a facet sequence which we refer to as a *pipe* $p = (e_0, \ldots, e_n)$. In fact, the initial facet element shown in response to search query represents already the first element of a pipe. To add an element to p, in our prototype, a set of controls including a plus-button appears below an element when hovering over it. Clicking on the plus-button adds a new pipe element above which then shows the relevant facets at this position of the pipe. To illustrate the concept, Figure 1 shows a pipe with five elements.

As a special feature of our prototype system, we decided to compute as term scores for a facet term t not the number of documents in D_q that are associated with t by default, but rather the number of distinct terms from the facets above and below that are connected with t through D_q. Though users are more familiar with document counts as term scores [17], we choose the concept of up- and downscores in our prototype as it is more expressive: By re-arranging the facets in the pipe, users can explore the relationships between arbitrary facet pairs. Document counts can always be obtained by inserting a facet with a 1:1 relationship to the document collection D as a neighboring element (such as the collapsed "Publications" facet in Figure 1).

Leaving the introduction of scoped facets to the next subsection, the term scores for a pair of neighboring pipe elements (e_{i-1}, e_i) are computed as follows: With all currently selected terms in the pipe elements $\{e_k \mid k = [0, i]\}$, a query is formulated and run against our system to retrieve the set of relevant documents $D_i \subseteq D$ at pipe position i. Following common practice, all selected terms of one pipe element are joined to an expression with disjunctions (`or`) in the query, while these expressions themselves are joined with conjunctions (`and`):

$$D_i = \{d | d \in \cap_{k=0}^{i} (\cup_{t \in e_k} I(t))\}.$$

I.e., D_i contains those documents that are associated with at least one selected term in every pipe element e_k. To compute the term scores for each term in e_i, it is counted how many distinct terms of e_{i-1} are related to the term through D_i:

$$score(t) = |\{t_j | t_j \in e_{i-1} \wedge \exists\, d : d \in D_i, I(t), I(t_j)\}|$$

The resulting counts are referred to as the *up-scores* of the terms in e_i. Canonically, the *down-scores* of the terms in e_{i-1} are computed by counting the number of distinct relations to terms in e_i. As shown in Figure 1, the up- and down-scores are displayed in the two leftmost columns of the child term table. Since the first and the last pipe-elements have only one neighbor, only one of the score types is defined, respectively.

Scoped Facets. With the concept described so far, the documents that can be explored at pipe element e_i are limited to the retrieved documents D_{i-1}, i.e., users can always only build subsets of the currently retrieved documents by adding a new element to the pipe. We argue that this limitation is undesirable, as users may want to begin an exploratory search with a very specific query, and then explore the document collection from there.

One possibility to overcome this limitation would be to only consider, for the computation of term scores at position i, the terms of e_{i-1} and e_i to formulate the respective search query. This solution, however, would trade the possibility of broadening a search by inserting a pipe element with the possibility to progressively narrow down the search results with more than one facet (e.g. in order to find the papers an author published at a specific conference).

Instead, we propose to give the user the possibility to specify the pipe elements which should be considered for the term scoring at each position of the pipe by introducing the concept of *scopes.* The scope s_i of a pipe element e_i is an integer in the range $[0, i]$. The scope determines which of the pipe elements below e_i, i.e., $\{e_k \mid k = [i - s, i]\}$, shall be taken into account for the computation of the term scores of e_i. By default, the scope is set to $s = i$, which means that all prior pipe elements are considered. In the special case of $s = 0$, the pipe element e_i is considered a new starting point, and no term relations are computed for (e_{i-1}, e_i). To set the scope of a facet in our prototype, the controls of our user interface feature a range slider (cf. second pipe element in Figure 1).

Though a simple concept, with the introduction of facet scopes, our system gains a powerful exploratory search feature which sets

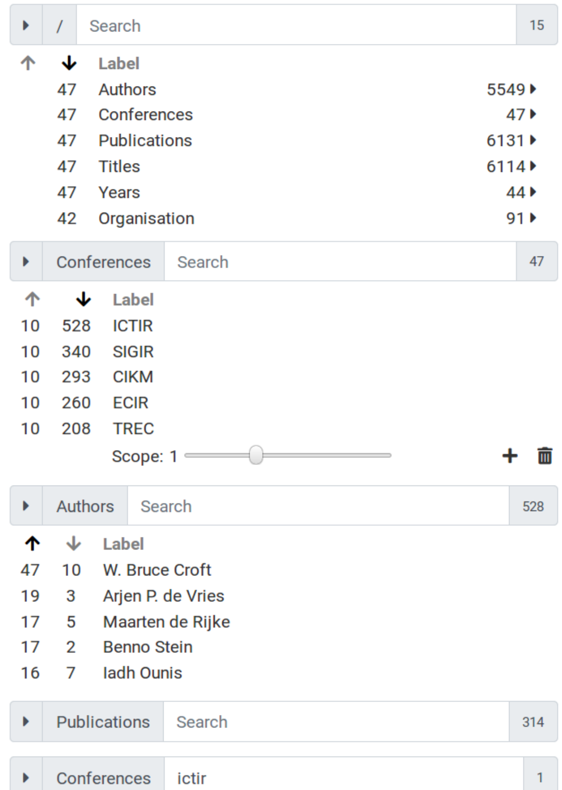

Figure 1: Screenshot of our exploratory search prototype highlighting the benefits of facet scopes. By adding the first facet (shown collapsed at the bottom) with a reduced scope to the pipe again, similar terms with respect to the intermediate facets can be determined.

it apart from prior work. A special usage pattern of facet scopes that we observe is that users apply a reduced scope to discover facet terms which are similar with respect to intermediate facets. For example, in Figure 1, the user discovers conferences similar to ICTIR in terms of author overlap.

4 EVALUATION

In order to evaluate our concept, we conducted a user study with 14 participants. As a baseline to compare our prototype system *podascope* with, we choose the digital library DBLP[2], which indexes a large dataset of currently around 4.5 million computer science research paper records. DBLP was chosen as our baseline system for two reasons: (1) The dataset indexed by DBLP is freely available, such that we could index the same data with our research prototpye. (2) The domain of DBLP matches the interests of our 14 participants (all researchers in various computer science disciplines), which is according to Kules and Capra a desirable circumstance for the evaluation of exploratory search systems [9].

In our user study, we were interested in two main questions: (1) Does our proposal of exploratory search pipes with scoped facets harm the search experience of users with simple search tasks that do not require query broadening? (2) Does our proposal increase

[2] https://dblp.uni-trier.de/

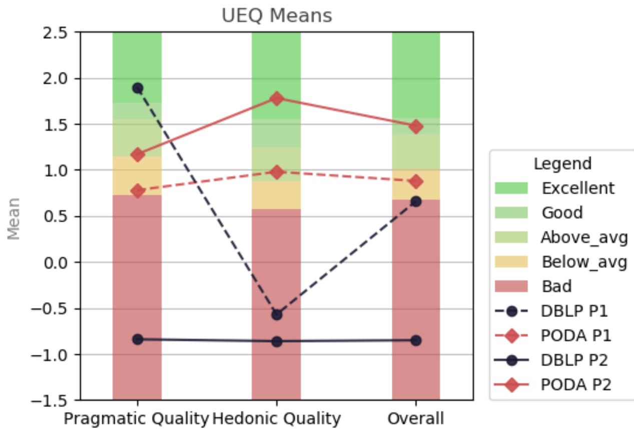

Figure 2: Results of the user study. The usability of both systems are assessed as average for simple search tasks (dashed lines). For complex search tasks, the usability of our prototype *podascope* (red) increases to good. For the baseline system (black) it decreases to bad.

the user experience for search tasks which are complex (do require query broadening)?

To answer these questions, the user study was divided into two parts. In the first part of the study, the participants were asked to solve five simple search tasks with both systems. During the study, an adviser was physically present and assisted the participants if needed. One half of the participants always started solving the task with the baseline system, whereas the other half started with our prototype. The five simple search tasks were (topics differed depending on the research domain of the participant):

- Find all papers with “exploratory search” in the title.
- How many authors published on this topic?
- Which author published the most papers on this topic?
- At which conferences did he/she publish these papers?
- What other papers did he/she publish?

After solving the simple search tasks, the participants were asked to evaluate the systems by answering the user experience questionnaire UEQ-S [3], which asks for four pragmatic and four hedonic quality dimensions. The results of this first evaluation is presented in Figure 2 (dashed lines). With an overall mean quality score of 0.88 (podascope) and 0.66 (DBLP), the user experience of both systems is assessed comparably well by the annotators, leading us to the conclusion that our concept did not harm the search experience for simple search tasks.

In the second part of the study, the participants were asked to solve the following three complex search tasks:

- Which five conferences are most similar to “ICTIR” in terms of author overlap?
- In how many years did each of these conferences take place?
- Who published most papers at these conferences?

To solve these tasks, the facet scope has to be reduced in our prototype system. For the baseline system, solving the tasks requires to first select all relevant facet terms of the intermediate query (e.g. all ICTIR authors), and then de-selecting the facets terms of the first query. Since DBLP does not allow to select multiple terms of a facet to formulate a disjunctive query, the first complex task could be solved only roughly with this system by selecting one of the authors. After working on the tasks, the participants were asked to answer the UEQ-S again (see solid lines in Figure 2): Whereas the baseline system is assessed worse than before (mean UEQ-S score of -0.85), the user experience score of our system increases to 1.48, leading us to the conclusion that our concept improves the exploratory search experience for the case of complex search tasks.

[3] https://www.ueq-online.org/

5 CONCLUSION

Our concept of exploratory search pipes with scoped facets contributes a novel and powerful exploratory search paradigm that facilitates the intuitive, continuous, and dynamic exploration of document collections. The user study conducted on the basis of a prototype of our concept in an exploratory search system revealed that for simple search tasks, our paradigm does not harm the overall search experience of users. For complex search tasks requiring the broadening of current search queries, the usability of our system is perceived as superior to a traditional exploratory search system.

REFERENCES

[1] Marcia J. Bates. 1986. Subject access in online catalogs: A design model. *Journal of the American Society for Information Science* 37, 6 (1986), 357–376.
[2] Tristan Blanc-Brude and Dominique L. Scapin. 2007. What do people recall about their documents? Implications for desktop search tools. In *Proc. of the 12th Int. Conf. on Intelligent User Interfaces, IUI, Honolulu, USA, 2007.* 102–111.
[3] Robert G. Capra and Gary Marchionini. 2008. The Relation Browser Tool for Faceted Exploratory Search. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '08).* ACM, New York, NY, USA, 420–420.
[4] Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. SearchLens: composing and capturing complex user interests for exploratorysearch. In *Proc. of the 24th Int. Conf. on Intelligent User Interfaces, IUI 2019, Marina del Ray, CA, USA, March 17-20, 2019.* 498–509.
[5] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. 1992. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In *Proceedings of the 15th Int. ACM Conf. on Research and Development in Information Retrieval (SIGIR '92).* ACM, New York, USA, 318–329.
[6] E. Garfield. 1970. When is a negative search result positive? *Essays of an Information Scientist 1* (1970), 117–118.
[7] Gene Golovchinsky, Abdigani Diriye, and Tony Dunnigan. 2012. The Future is in the Past: Designing for Exploratory Search. In *Proceedings of the 4th Information Interaction in Context Symposium (IIIX '12).* ACM, New York, NY, USA, 52–61.
[8] Philipp Heim, Thomas Ertl, and Jürgen Ziegler. 2010. Facet Graphs: Complex Semantic Querying Made Easy. In *The Semantic Web: Research and Applications.* Springer Berlin Heidelberg, Berlin, Heidelberg, 288–302.
[9] Bill Kules and Robert Capra. 2008. Creating exploratory tasks for a faceted search interface. In *Proceedings of the Second Workshop on Human-Computer Interaction.*
[10] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46.
[11] F. Moretti. 2013. *Distant Reading.* Verso.
[12] G.M. Sacco and Y. Tzitzikas. 2009. *Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience.* Springer.
[13] m.c. schraefel, Max Wilson, Alistair Russell, and Daniel A. Smith. 2006. mSpace: Improving Information Access to Multimedia Domains with Multimodal Exploratory Search. *Commun. ACM* 49, 4 (April 2006), 47–49.
[14] Elaine Toms. 2000. Serendipitous Information Retrieval. In *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries.* 11–12.
[15] Yannis Tzitzikas, Nikos Manolis, and Panagiotis Papadakos. 2017. Faceted exploration of RDF/S datasets: a survey. *Journal of Intelligent Information Systems* 48, 2 (01 Apr 2017), 329–364.
[16] Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query-Response Paradigm.* Morgan & Claypool.
[17] Max L. Wilson and m.c. schraefel. 2006. mSpace: What do Numbers and Totals Mean in a Flexible Semantic Browser. (2006).
[18] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted Metadata for Image Search and Browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03).* ACM, New York, 401–408.