A Study on User-Controllable Social Exploratory Search

Cecilia di Sciascio

Know-Center GmbH, Graz University of Technology

cdisciascio@knowcenter.at

ABSTRACT

Information-seeking tasks with learning or investigative purposes are usually referred to as exploratory search. Exploratory search unfolds as a dynamic process where the user, amidst navigation, trial-and-error and on-the-fly selections, gathers and organizes information (resources). A range of innovative interfaces with increased user control have been developed to support exploratory search process. In this work we present our attempt to increase the power of exploratory search interfaces by using ideas of *social search*, i.e., leveraging information left by past users of information systems. Social search technologies are highly popular nowadays, especially for improving ranking. However, current approaches to social ranking do not allow users to decide to what extent *social* information should be taken into account for result ranking. This paper presents an interface that integrates social search functionality into an exploratory search system in a user-controlled way that is consistent with the nature of exploratory search. The interface incorporates control features that allow the user to (i) express information needs by selecting keywords and (ii) to express preferences for incorporating *social wisdom* based on tag matching and user similarity. The interface promotes search transparency through color-coded stacked bars and rich tooltips. In an online study investigating system accuracy and subjective aspects with a structural model we found that, when users actively interacted with all its control features, the hybrid system outperformed a baseline content-based-only tool and users were more satisfied.

KEYWORDS

exploratory search; social search; information filtering; preference elicitation, hybrid ranking

1 INTRODUCTION

Learning and investigative tasks entail a combination of querying and browsing actions occurring as part of a dynamic process often called exploratory search [27]. Acquiring knowledge about a new topic is rarely fulfilled with a single query. Conversely, each bit of new knowledge triggers changes in information needs. In other words, learning is the result of a discovery cycle of several queries intermingled with the analysis of retrieved resources, where facts from a large volume of information fall in a conceptual representation [33].

It has been long acknowledged that exploratory search process can benefit from the "collective wisdom" of many people who can explicitly or implicitly collaborate in the search process [32]. Yet, almost all work on collaborative exploratory search focuses on explicit synchronous collaboration context [32, 41] ,with almost no studies of exploratory search interfaces based on "implicit collaboration" where future users can leverage information left by past users of information systems. This forms a sharp contrast with

general stream of research on information retrieval where "implicit collaboration" known also as social search [4, 16] has emerged into a broad stream of research. In this paper we attempt to bridge this gap by investigating the value of two social search approaches (one based on the use of social tags and another on collaborative user matching) in the context of exploratory search.

Our approach to support social exploratory search is to allow the user to filter documents by controlling the influence of search terms, as well to fuse traditional query-based document relevance with relevance produced by two "social" sources. To visualize this multi-source relevance, our interface uses visual cues to augment the ranked list, whereby relevance is not represented as a single score, but instead as the contribution of multiple dimensions, i.e. search terms and social sources. Intuitively, adding on control and transparency at multiple levels could possibly turn out into an overly complex system. Hence, our interest is to unveil whether interacting with a social exploratory search system based on multi-source relevance, like the one here proposed, can yield measurable benefits for users conducting exploratory search, in contrast to other simpler (but still interactive) systems.

In summary, the contributions of this paper are: (i) an interactive exploratory search system, with a user interface designed to maximize transparency and controllability at multiple dimensions (search terms and relevance sources); (ii) a hybrid social ranking algorithm that accounts for fast changes in user needs and produces decomposable scores to facilitate representations in the UI; and (iii) a user study revealing that active users that control the system obtain more accurate results, are able to appreciate transparency features and are more satisfied.

2 RELATED WORK

2.1 Exploratory Search

Information seeking is a widely studied phenomenon [20, 36], as finding and organizing pieces of information occupies a large portion of our daily productive time. Information retrieval (IR) systems have grown as the preferred solution for contextualized search due to their ability to narrow down the number of entries that need to be inspected at a time. However, this kind of system requires precise user-generated queries. As the user learns about certain topic, queries have to be iteratively reformulated to express evolving information needs. Formulating queries has proven to be more complicated for humans than plainly recognizing information in a visual manner [20], which is why the combination of IR with HCI techniques has led towards a shift in the way users search. Browsing search strategies, which rely on on-the-fly selections, navigation and trial-and-error, are associated with the term that Marchionini et al. [27] coined "exploratory search". By definition, exploratory search is open-ended, i.e. the user starts the search

with some initial query in mind and discovers the next query terms along the way, as she finds and scrapes new information.

Over the last decades, several approaches attempted to foster deeper exploration and sensemaking of search results or large document collections. Ranked lists have prevailed as the paradigm for presenting results, due to their familiar format and because users know where to start inspecting items. Although ranked lists alone are regarded as opaque and under-informative [19], Shani et al. [35] suggest that: (i) explaining relevance scores encourages users to explore beyond the first two results, and (ii) users prefer bars over numbers or the absence of graphical explanations. In addition, the use of visualizations has been favored due to their capacity to convey document relevance by exploiting pre-attentive patterns. For example, tile bars [19] encode relative frequency of query terms with compact shaded blocks, whereas other approaches complement lists with visual metaphors [29], similarity-preserving layouts [17] or POI-based visualizations [30].

The trend in modern search systems seems to go in the direction of highly interactive UIs supported by ever-growing artificial intelligence methods. A few examples include task models [3], categorized overviews [39], adaptive visualizations [1] and intention modeling [34]. In this context, our work follows a recent trend in the area of interactive UIs for exploratory search, which extends the popular faceted search approach [40] to a context with no hierarchical metadata. In this group of UIs, the original query is used to produce not only a list of results, but also a list of most important information entities covered by the retrieved results - from simple unigrams [12, 22] to named entities [2] and user-specified keywords [34]. These entities uncover the *aboutness* of the results and serve as interactors for further exploration of the generated ranked list.

More specifically about the uRank interactive ranking approach [12], we replicate features which allow the user to manipulate the weights of the selected entities to adjust the original ranking to their emerging needs. This system extracts keywords from titles and abstracts and displays them in the UI as interactors. As the user selects (or types) keywords of interest, a document ranking visualization is re-sorted in real-time, thus promoting a *search-by-browsing* information access paradigm. di Sciascio et al. found that search-by-browsing supported by this kind of adaptive system incurs in lower cognitive load without hindering user performance, compared to conventional list-based UIs [13].

After several users have conducted exploratory search with adaptive systems, they leave traces behind that provide hints as to what combinations of terms were fruitful in their searches, or which documents were preferred by users with similar information needs. The availability of this kind of traces paves the way to support future users by complementing exploratory search with social wisdom.

2.2 Social Search and Interactive Fusion of Relevance Sources

Social search is common name for a group of information retrieval approaches that use traces left by past users of information systems to help future users in the search process [15, 16]. The traces of past users can be obtained from multiple sources, such as search systems logs, social tagging systems, Web site logs, and various

social media systems. In turn, this information could be obtained to assist users in different stages of their search including from query formulation [6], matching and ranking of results [7], and augmentation of results [4]. Most popular target of social search techniques is ranking of search results. The idea of social ranking is to combine traditional content-based relevance of search results with relevance measures obtained from social sources. For example, documents that are frequently selected in response to similar queries as well as results that have been recognized by target users by bookmarking, tweeting, or other kinds of sharing in social systems have high social relevance and have to be promoted in the list of search results. Traditionally, the fusion of content-based and social relevance is done automatically by learning the weights of different sources using learning-to-rank or similar data-driven approaches [23]. The automatic fusion, however, does not fit well to the nature of exploratory search where user might want to decide to what extent social information should be considered to rank results of a particular query. In our paper, we suggest an interactive controllable fusion of content-based and social relevance judgment and present an interface that allows users to control this fusion in a way that is consistent with modern information exploration interfaces.

Interactive fusion of relevance sources is not an entirely new topic in the field of information access and intelligent interfaces. The majority of research on this topic, however, has been performed in the area of hybrid recommender systems (RS) [9]. In the context of so-called parallel hybridization [9], a recommender system has to fuse relevance judgment obtained from difference sources or approaches. Just like in case of social search, the traditional approach to source fusion in hybrid system is automatic, where the influence of each component is determined using some machine learning approach and stays the same. This, however, does not allow for accommodating to real-life context where the importance of each source could depend on the varying user needs. For example, a movie recommendation could be a hybrid of personal collaborative filtering and social recommendation collected through social connections. When watching alone, a user may put more emphasis on the personal part, while when selecting a movie to watch with friends, the social part should be more valuable. To offer more flexibility, for example, *TasteWeights* [8] generates music recommendations from multiple sources and presents them in a visually rich interface that allows the user to interactively change the contribution of each individual source to the ranked list of results. *SetFusion* [31] allows to control the contribution of three recommender sources in a hybrid RS for research talks and indicates which sources were used to recommend each item with color-coding. User-controlled source fusion is considered to be a part of a broader stream of work on controllability for RS. A range of user studies on interactive and controllable RS indicates that when users are given the chance, they do make use of their ability to control the system [14] and are more satisfied [8]. For further details, He et al. [18] elaborated a comprehensive survey of interactive RS, including visualization, presentation and interaction aspects.

Within the area of information retrieval, the idea of interactive fusion attracted much less attention, since the presence of several

independent relevance sources is less typical. Also, extensive exploration with an interactive source-fusion interface is not natural in most search contexts. One exception is the work on personalized search. In this context, each search result has two relevance scores - relevance to the query and relevance to the user profile. Traditionally, personalized search systems perform automatic fusion of query-based and profile-based relevance rating to offer personalized ranking [28]. However, Ahn et al. [3] demonstrated that allowing users to choose how to combine these two ratings for each individual query can increase all performance aspects.

Social exploratory search offers an excellent opportunity to apply interactive fusion of relevance sources outside of the traditional recommendation context. On one hand, the social search approach provides several relevance sources beyond regular query-based relevance (each type of socially collected information could be used for independent relevance ranking). On the other hand, exploratory context makes interactive exploration of results with user engagement into ranking quite natural. In this paper we investigate the value of user controllable ranking to combine a traditional query-based relevance with two "social" relevance sources - one based on user-assigned tags, and one based on collaborative matching between the current and past users. The next sections describe control and transparency features implemented in the UI and the ranking model supporting the user task.

3 SYSTEM OVERVIEW

In order to blend exploratory and social search, we replicate features of an adaptive system designed for exploratory search (ES) of textual documents [13]. The basic system promotes a search-by-browsing information access paradigm, using keywords extracted form search results as interactors to refine a document ranking upon evolving information needs.

As the user interact with extracted keywords and bookmarks resources, our system can learn about the importance of certain keywords and their connection to documents. As a result, the social-enhanced ES system incorporates collaborative and tag-based filtering methods into a hybrid ranking model, whereby (i) tagging data used for training purposes is implicitly generated from book-marking behavior, and (ii) the interface allows the user to weigh the fusion between content information and *social wisdom* in the resulting ranking.

In the remainder of this section we briefly describe the features of the user interface (Section 3.1) and the elements of the system that enable social search functionalities alongside exploratory search.

3.1 The User Interface

Exploratory search is often motivated by a complex information problem where the user has little understanding of terminology or information space structure [38]. Our tool supports the information seeking process with features for exploration and explanation, which largely correspond to control and transparency in the UI. The motivation for control and transparency features comes from research on user experience with recommender systems, whereby users can gain in confidence and performance when the system does not behave like a black box but is instead flexible to users' preferences [8] and provides sufficient explanations [21].

As a starting point for exploration, the 12 most frequent keywords appear arranged in an inline fashion, at the top of the UI in the *Keyword Box* (Figure 1.A). By hovering on a first-row keyword, the user can discover other keywords frequently appearing together in the documents' text, which appear on the second row (connected by tree-shaped links).

3.1.1 Controlling the System. The *Query Box* (Figure 1.B) is the component that captures current user's information needs, expressed in terms of iterative interactions with keywords. Possible interactions inherited from the original system include keyword addition (either by selection ot manually typing), deletion and weighting. We extended these options with filtering functionalities, so that users can also: (i) filter individual keywords, so that bearing documents remain in focus and the rest are dimmed (▼) button visible on mouse over); and (ii) enable the "all keywords" filter (toggle button on the right side of the *Query Box*). The latter switches to a strict bearing criterion (from "or" to "and"), so that the ranking includes only documents where all selected keywords appear in the text. Selected terms in the *Query Box* appear joint in a single body, as shown in Figure 2.

The main addition in the UI is the possibility for the user to tell the hybrid model the desired impact of content-based and social relevance on the joint ranking. Sliders in Figure 1.C adjust the weights of the three methods in the hybrid ranking model, namely:

📄 Content: It ranks documents based on terms contained in their titles and abstracts (Section 3.3.1).

🏷 Tags: based on selected keywords matching tags associated to documents by bookmarking actions of past users (Section 3.3.2).

👥 Collaborative: brings documents previously bookmarked by other users with similar interests (Section 3.3.3)

3.1.2 Transparency in the UI. A highly controllable UI could be cumbersome if users cannot perceive the effect of their actions. Therefore, we rely on graphical explanations to convey system decisions and mitigate the complexity of multiple control features. Taking into account that the tool allows for tuning both keywords and model parameters, we had limited options for the use of color to convey score contributions. We decided to use a categorical color palette to represent keywords and thus maximize the amount of information encoded with color. Color-coded keywords are visual cues that pop out from their surroundings, enabling the user to pre-attentively recognize them in the text and perceive their general context prior to conscious reading.

Color-coded stacked bars embedded within documents in the *Document Ranking* (Figure 1.D) indicate the scores computed thereof. The overall width is the total document score, while individual bars in the stack denote the contribution of each keyword to the total score. This means that the score produced by each method in the ranking model is broken down into its constituent keyword-based sub-scores. Sub-scores are then added keyword-wise to represent a single bar in the stack. For example, in Figure 3 the length of the green bar is obtained as the addition of the content, tag and collaborative sub-scores for the keyword "tool".

In turn, icon hints at the bottom-right corner of individual items in the *Document Ranking* reveal at a glance which methods ranked a given document. The icons match the labels in the model parameter
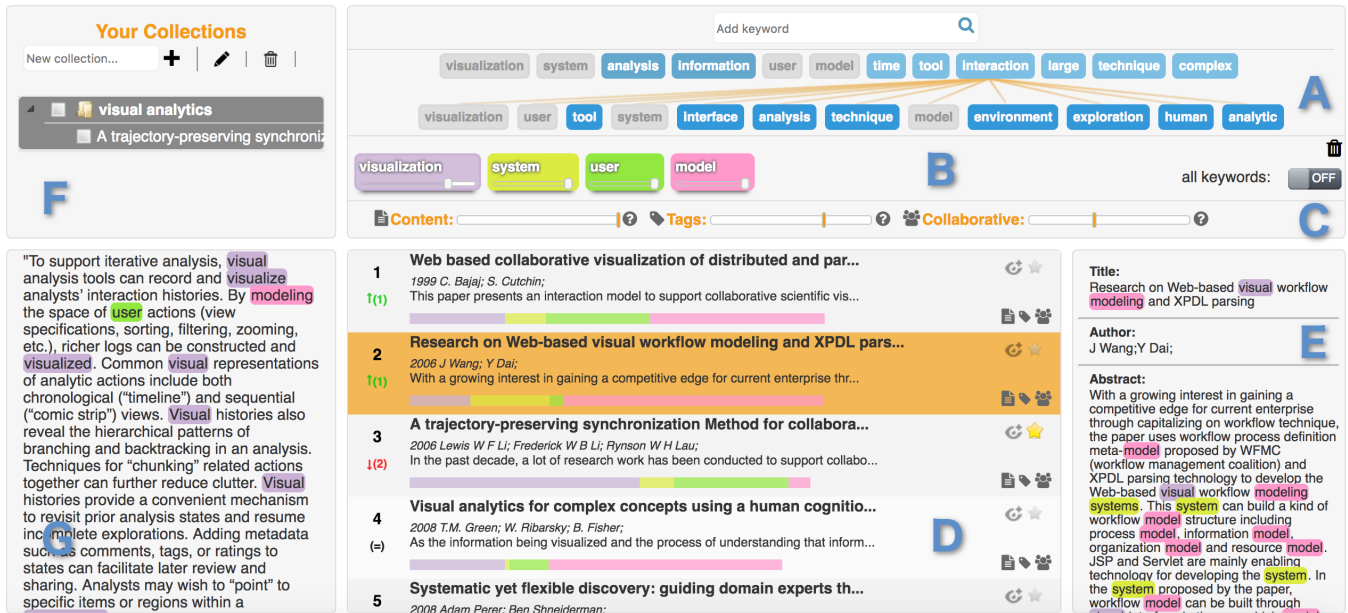
Figure 1: The UI of the exploratory search system provides features to control a hybrid recommender system and graphic explanations of the recommendation provenance.
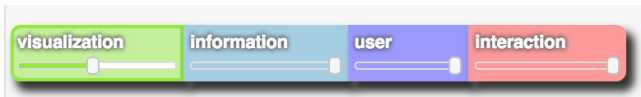

Figure 2: Switching on the "All Keywords" filter sets a strict "and" constraint. Tags in the *Query Box* appear grouped.

controls (Figure 1.C). Visualizing the overlap between keywords and information source is possible by hovering over the stacked bars. The tooltip in Figure 3 explains which keywords appear in the document, which ones have been used to tag the document and which ones are popular among users with similar interests.

When the user clicks on a list item, the *DocumentInfo Panel* displays the available text and metadata information for the selected document. Terms matching selected keywords in the *Query Box* are highlighted in matching colors, enabling the user to readily spot them in the text prior to conscious reading. The same principle is applied for the notepad in Figure 1.G. As the user saves a document into a collection, this piece of text is augmented with colored terms, highlighting the parts of the text related to bookmarked resources.

3.2 Generating Tagging Data

Text-mining or topic-modeling methods generate machine-based content descriptions, often referred to as keywords or key phrases (in the case of unigrams and n-grams, respectively). Conversely, tags are regarded as human-generated content descriptors given by one or more users. Social tagging data is the space where relationships between users, resources and tags are captured.

Unlike a typical social tagging system that serves as a platform for users to tag resources, our system is designed for users to search
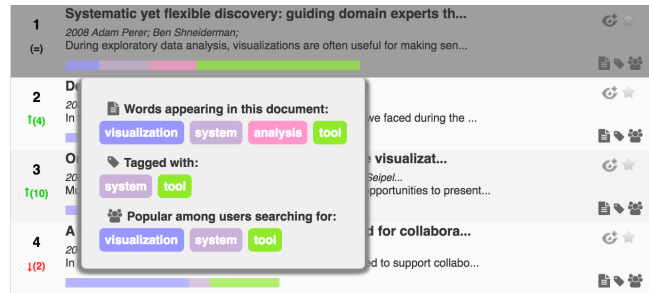

Figure 3: Hovering over a document's stacked bars shows a tooltip informing which tags influenced each RS.

and explore documents. Hence, our approach to incorporate social information does not work with social tagging data in the strict sense, as users do not explicitly assign tags to documents at any time. Instead, we leverage bookmarking behavior to infer human-approved content descriptors, which we treat as (so-called) tags, i.e. tags are not generated but rather approved by humans.

As the user explores search results by selecting extracted keywords from the *Keyword Box*, they save relevant documents along the way as bookmarks. Bookmarking actions provide useful information in a twofold manner: *i)* the user establishes an explicit preference for a given document, and *ii)* the user implicitly considers that the document is relevant for their current information needs, which are expressed by the selected terms present in the *Query Box* at the moment of the bookmark event. We can assume that a user regards the currently selected terms as good content descriptors for the bookmarked document, and hence refer to them as tags in our tagging space. One reason to treat human-approved

content descriptors as tags is that humans largely tend to generate keyphrases that already appear in the text [11]. For clarification, in the remainder of this paper we refer to keywords or terms as automatically extracted unigrams, and recall tags as terms used to implicitly tag bookmarked documents. Therefore, a document is tagged with *tags* $t_1, \cdots, t_n$ if a user bookmarks the document after selecting a set of *terms* $t_1, \cdots, t_n$ in the UI.

Formally, B is the set of all bookmarks, where a single bookmark is represented by a tuple of the form $(u, d, Q) \in B$. Thus, a bookmark entry denotes that user u saved a document d and implicitly tagged it with selected terms $Q = \{t_1, \cdots, t_n\}$. Hereby, we define the folksonomy $\mathcal{F} = \langle U, D, T, A \rangle$, where U is the set of all users, D the set of all documents, T the set of all tags and A (Formula 1) is a function mapping a bookmark $b \in B$ into a set of $|Q|$ triples of the three entity types, (u, d, t).

$$A = \{ (u, d, t) \mid u \in U, \ d \in D, \ t \in Q \subseteq T, \ \exists \ (u, d, Q) \in B \} \quad (1)$$

Triples mapped by A conform the tagging space that represents ternary relationships between users, documents and tags. We break down ternary relationships into three matrices as follows:

- $R \in \mathbb{N}^{|U| \times |D|}$ is similar to the user-item matrix employed in CF. Unlike rating data, where values typically range from 1 to 5, in this context user preference for an item is binary, thus $r_{u,d} = 1$ if user u bookmarked document d, 0 otherwise.
- $P \in \mathbb{N}^{|D| \times |T|}$ is the document-tag matrix, such that the ith row represents the profile vector for the ith document, in terms of tags associated to it at the moment it was bookmarked.
- $M \in \mathbb{N}^{|U| \times |T|}$ is the user-tag matrix, such that $m_{u,t}$ indicates the number of documents bookmarked by user u after selecting term t (and implicitly using t as a tag).

In a nutshell, we extract explicit and implicit information contained in bookmarking behavior to generate tagging data, which is then employed to train the social-based ranking methods described in the next section.

3.3 Ranking Model

As user u adds a set of terms $Q = \{t_1, \cdots t_n\}$ to the *Query Box*, the system computes independent document scores with three different models and ranks them according to an overall hybrid score. The content-based model (CB) (Section 3.3.1) produces scores based on textual information from documents' titles and abstracts. This is the ranking method built in the original adaptive search system. The tag-based (TB) and collaborative models (Sections 3.3.2 and 3.3.3) are incorporated to extend the original system and support social exploratory search.

Interaction and presentation were important factors steering the design and implementation of these methods, namely:

(1) For the system to adapt to sudden changes in information needs and provide a sense of real-time response, each method has to compute new document scores on-demand.
(2) Then for presentation purposes, they should generate decomposable scores that can be graphically explained, i.e. calculate separate scores for each selected term $t \in Q$.

In this section we describe the three models and the hybridization of the final score.

3.3.1 Content-based (CB). It applies IR techniques to rank documents by the relative frequency at which the selected tags (keywords) appear in titles and abstracts. We build a vector space model and compute document-query similarity as the weighted cosine similarity (Formula 2).

$$s_{cb}(d, Q) = \frac{1}{|Q|} \sum_{t \in Q} \frac{tfidf(t, d) \ \cdot \ w_{u,t} \cdot \Gamma(d, Q)}{\| d \| \cdot \| Q \|} \quad (2)$$

, where $tfidf(t, d)$ is the tf-idf score for term t in document d, $w_{u,t}$ is the weight assigned to term t by user u in the UI, while $\| d \|$ and $\| Q \|$ are the Euclidean norms for vectors d and Q, respectively. Lastly, $\Gamma(d, Q) = exp(\alpha \ (|d \cap Q| - |Q|))$ is a decay function that penalizes documents not containing all selected keywords. We set the α parameter to 0.25 to soften the decay rate.

3.3.2 Tag-based (TB). The tag-based model measures the strength between the target user's selected terms Q and a document d in the social tagging space. In other words, this model expresses to what extent terms Q are good content descriptors of d from the perspective of other users that bookmarked (and implicitly tagged) d in the past.

Given the document-tag matrix P, the similarity between term t and document d, $sim(t, d)$ is computed as the conditional probability of d to be bookmarked given t, which in Formula 3 is represented as the ratio between the tagging frequency of d under t and the tagging frequency for t across all documents. s_{tb} is then the probability of t and d occurring together, i.e. $p(t) \cdot p(d \mid t) = p(t \cap d)$.

$$s_{tb}(u, d, Q) = \sum_{t \in Q} w_{u,t} \cdot \frac{p_{t,d}}{\sum\limits_{d' \in D} p_{t,d'}} \quad (3)$$

3.3.3 Collaborative. The nature of this model resembles that of a collaborative filtering recommender, except that traditional CF was conceived to learn a model from rating data, whereas ratings in the case of bookmarks is binary. i.e. a document is either bookmarked or not. The UB model estimates the likelihood for a document d to be bookmarked by target user u based on the strength between u's neighborhood V and d. Since u's tag preferences are frequently updated through the UI, similarity between two users is in this case agnostic of u's past search interests. Instead, v is similar to u if v has bookmarked any document with selected terms Q in the past.

Given the user-tag matrix M, we calculate the similarity between u and v as the weighted ratio between the times v used t and the total times t was used by any user to bookmark any document, as denoted in Formula 4.

$$sim(u, v) = \sum_{t \in Q} w_{u,t} \cdot \frac{m_{v,t}}{\sum\limits_{v' \in V} m_{v',t}} \quad (4)$$

The neighborhood for u, V is formed as the union set of the neighborhood for each $t \in Q$, V_t, as depicted in Formula 5, such that, V_t is the set of users that implicitly tagged any document with term t. Then, V is trimmed to the k top neighbors ($k = 10$ by default).

$$V = \bigcup_{t \in Q} V_t \mid v \in V_t \; if \; m_{t,v}^{\mathsf{T}} > 0 \tag{5}$$

We extend the user-item matrix R into matrix $\tilde{R} \in \mathbb{R}^{|U| \times |D|}$, such that the preference of user v for document d is 1 if v bookmarked d, $sim_J(v, d)$ otherwise, where $sim_J(v, d)$ is the Jaccard coefficient between v's user profile and the d's document vector. The final s_{ub} score is obtained as shown in Formula 6 by averaging the product of user-neighbor and neighbor-document similarities across all neighbors in V.

$$s_{ub}(u, Q, d) = \frac{1}{|V|} \sum_{v \in V} sim(u, v) \cdot r_{v,d} \tag{6}$$

3.3.4 Hybrid Overall Score. Relative ranking weights can be interactively adjusted in the $[0, 1]$ range and are represented by vector W. Individual weights are balanced with the softmax function, so that $\sum_{i \in |W|} W_i = 1$. The outputs of the three models are first min-max normalized to avoid that higher scores from one method undermine the contribution of the others. Then, the hybrid score for d is the linear combination shown in Formula 7, where $\hat{s}_{cb}$, $\hat{s}_{tb}$ and $\hat{s}_{ub}$ are the normalized values for the corresponding scores.

$$s(u, Q, d) = W_{cb}\, \hat{s}_{cb}(d, Q) + W_{tb}\, \hat{s}_{tb}(u, d, Q) + W_{ub}\, \hat{s}_{ub}(u, d, Q) \tag{7}$$

To allow for explanations of individual keyword contributions, all methods calculate separate scores for each $t \in Q$, which are then fused and represented as color-coded bars in the UI. In the next section we analyze objective and subjective implications of fusing content, tag-based and collaborative relevance scores.

4 USER STUDY

We conducted an online user study to assess the worth of *user-controllable social exploratory search* (SES) in contrast to pure ES. Note that we do not aim to compare ES with traditional list-based UIs (c.f., [13]), instead we assess social exploratory search when users are aware and able to control it, compared to plain individual ES. While developing the hybrid tool with numerous controllable and explanatory features, we hypothesized that these additions would positively influence system accuracy and user experience. However, a large amout of features add complexity to the UI, which could results in a system that is too difficult to use and understand. In this study we address such concerns in detail.

4.1 Evaluation Methodology

The study used a between-subject design with conditions ES and SES. As the social search system requires training data to produce document scores, we split the study execution in two stages: (i) with baseline ES condition, (ii) with hybrid SES condition, after training social models with bookmarks collected in (i). Participants in both ES and SES performed the same task. Figure 4 provides an overview of the steps followed in the study execution.

4.1.1 Baseline System. For the ES condition, we opted for a system that already supports exploratory search, but lacks social search capabilities. We used the base version of our system with
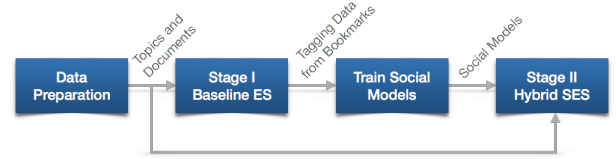

Figure 4: Stages in User Study

the ranking computed solely on the CB model. The interface in ES resembles that of the hybrid system, as shown in Figure 5, except for the absence of sliders to control the hybrid model and explanatory hints and tooltip described in Section 3.1.2.

As a result the baseline system includes most features of the hybrid system, but is inherently less complex. In other words, the user only needs to care about selecting keywords to refine information needs, without concerning about model parameters.
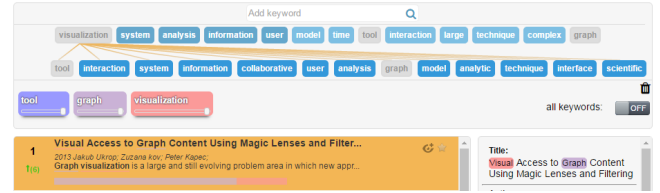

Figure 5: UI of baseline system employed in the user study. Documents are ranked solely by the CB method.

4.1.2 Data Preparation. The need to collect training data imposed a constraint on the number of topics included in the study. Therefore we chose 5 topics from the Computer Science field: *Augmented Reality, Visual Analytics, Recommender Systems, Deep Learning* and *Human-Computer Interaction*. With the help of three experts with research experience in the chosen topics, we prepared topic descriptions by selecting 2 or 3 paragraphs from known literature and removed all references from the text. We then generated datasets with results from a script that performed several queries to Mendeley's API[1] and manually added other documents suggested by the experts. Each dataset contained over 800 document surrogates.

After concluding the first stage, we had collected 1395 bookmarks. We further enlarged the training pool by asking the experts to perform the same tasks as regular participants, though without filling the post-task survey. Before starting the second stage, we generated tagging data as explained in Section 3.2 and trained the tag-based and collaborative models.

4.1.3 Task and Procedure. First, participants had to complete a step-by-step tutorial introducing the main features of the system. Participants had to perform interactions such as keyword selection and weighting, adjusting ranking model parameters, bookmarking and collection management. The tutorial also covered explanatory features in the UI such as color-coded bars and ranking-type icons (only in hybrid UI).
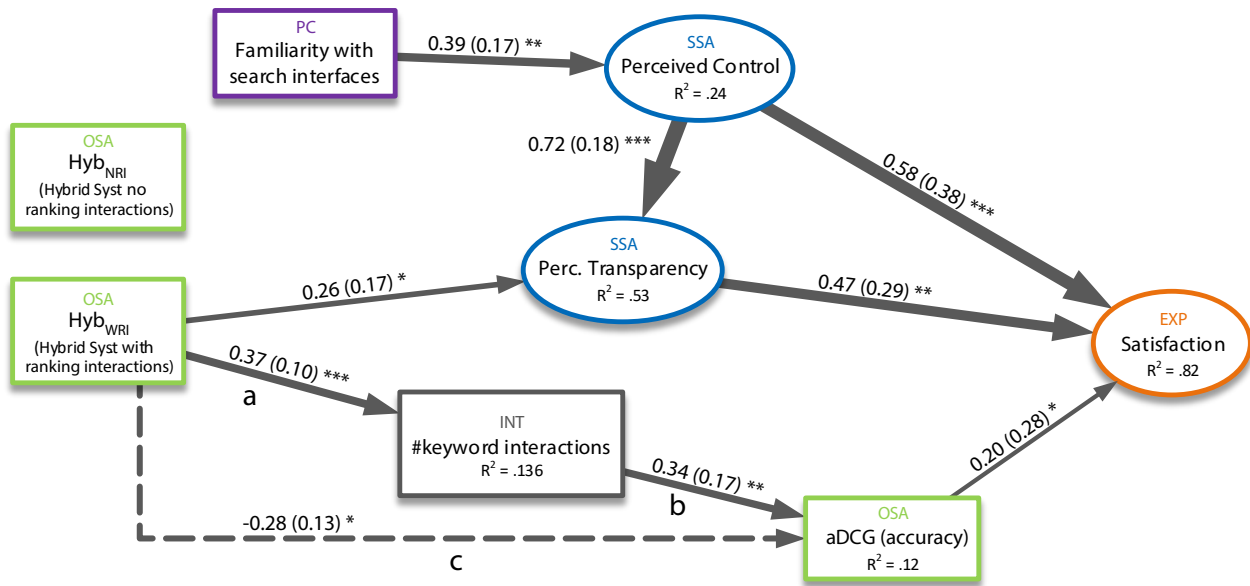
[1]http://dev.mendeley.com/

Figure 6: Structural model of system accuracy and user experience.

Elements of Evaluation Framework [26]: Objective (*OSA*) and Subjective System Aspects (*SSA*), Personal Characteristics (*PC*), Interactions (*INT*) and Experience (*EXP*). Pathways represent significant causal relationships. Numbers and thickness indicate regression coefficients (with robust SE). Factors are scaled with variances set to 1. Full standardized solution is reported (both latent and observed variables are standardized), so that path coefficients indicate increments in SD units. Full (dashed) lines indicate significant positive (negative) coefficients.

Significance: *** $p < .001$, ** $p < .01$, * $p < .05$.

After the tutorial, participants were presented with a view prompting them to sort the five topics by familiarity. The least familiar topic (i.e. at the bottom) was automatically picked. The task consisted in organizing and collecting resources for the given topic. We asked participants to imagine they had to write an essay about the assigned topic. We provided the corresponding topic description, which served as the introduction section. Participants had to read it and find relevant concepts that helped them define subsequent sections of the fictitious essay. Then they had to work with the assigned tool and find relevant documents for each section. To fulfill the task, participants had to create at least 3 collections (one per section) and bookmark at least 5 documents per collection, i.e. at least 15 in total. Upon completion, the participant had to click on "Done" and fill a survey for subjective feedback.

4.1.4 Participants. We recruited participants through the crowd-sourced platform Prolific[2] and from mailing lists of colleagues in the field. In the case of Prolific, we set as pre-screening eligibility requirement that candidates had to hold at least a Bachelor degree in Computer Science. Since the tasks required dealing with scientific content, it was crucial to address users with the appropriate background. As for the field of expertise, the intention was to match the topics covered in the prepared datasets. A total of 79 people took part in the study, from which 43 worked with the baseline system and the remaining 36 with the hybrid UI. 10 participants were recruited via e-mail, while the other 69 completed the study on Prolific. Demographic information is summarized as follows:

- *Age*: 1 [< 20] years old, 66 [20 − 29], 12 [30 − 39].

- *Gender*: 62 male, 17 female.
- *Country of birth*: India (18), USA (15), UK (9), other (37).
- *Highest level of education*: 39 Bachelor (BSc), 23 Master (MSc or similar), 17 other.
- *Proficiency in English language*: 6 basic, 9 intermediate, 30 advanced, 34 native speaker.
- *Familiarity with search interfaces, e.g. Google* (−3 = not at all familiar, 3 = very familiar): *mean* = 2.65, *sd* = 0.78.

4.1.5 Measurements.

Logged data. Both systems logged user behavior in terms of keyword manipulations, document clicks, bookmarks, and for the hybrid tool, manipulations of model parameter sliders. We assigned timestamps to every action and recorded the position of bookmarked items in the list.

System Accuracy. We computed *Discounted Cumulative Gain* (*DCG*) for each document d, defined as $DCG_d = \frac{2^{rel_d}-1}{\log(rank_d+1)} \cdot rel_d$ can take either 1 (if bookmarked) or 0 (otherwise). We then obtained the mean value per user, *aDCG*, by averaging *DCG* scores across all bookmarks performed by a single participant.

UX Scale for Subjective Feedback. We created a user experience (UX) scale adjusted to our task scenario, instead of using a standard but too generic questionnaire, e.g. SUS. Latent constructs, e.g. "satisfaction", cannot be directly measured and are instead inferred from a number of observed indicators, e.g. survey questions. The survey comprised 28 questions covering different UX dimensions. Answers were recorded on a 7-point Likert scale (-3 = strongly disagree, 3

= strongly agree), with some questions worded in negative tone (reverted for the analysis). We therefore built the model shown in Figure 6 in a two-step approach [5]: first we fitted a measurement model (to calculate factor loadings) and then a structural model (measurement model plus causal relationships, see Section 4.1.6). UX concepts and directionality of causal effects are partly grounded on the framework proposed by Knijnenburg et al. [26] to model the user experience with recommender systems.

We first planned for 5 latent constructs: 3 subjective system aspects (SSA) (perceived control, transparency and result quality) and 2 experience (EXP) (choice satisfaction and satisfaction with the system). "Choice satisfaction" and "satisfaction with the system" had poor discriminant validity (inter-factor correlation = .97), which means that users did not distinguish them from one another and perceived the two constructs as the same one. Since they both measure "satisfaction" aspects, we merged them into a single factor[3] and named it as such. In total 8 survey items were excluded from the analysis due to low communality or high cross-loading. In particular, all 4 items for "perceived result quality" had to be excluded because 2 of them had low communality[4].

The final three factors listed below showed good convergent validity (AVE), internal consistency reliability (Cronbach's α)[5] and discriminant validity[6].

- *Perceived Control* (SSA) ($AVE = .62$, $\alpha = .82$): 3 items, e.g. "The system allowed me to refine my search terms easily"
- *Perceived Transparency* (SSA) ($AVE = .62$, $\alpha = .82$) 3 items, e.g. "The system gave me a sense of transparency"
- *Satisfaction* (EXP) ($AVE = .59$, $\alpha = .93$) 9 items, e.g. "I'm satisfied with the documents I bookmarked", "I see a potential benefit in using the system for research tasks".
- Also, the question "Are you familiar with search user interfaces?" is a single personal characteristic (PC) observed indicator.

Confirmatory factor analysis with a mean- and variance-adjusted weighted least squares estimator yielded admissible model fit[7], $\chi^2(87) = 105.68$, $p = .08$; $CFI = .994$; $TLI = .993$; $RMSEA = .051$ with $90\% CI = [0, .083]$.

4.1.6 Hypotheses in Structural Model. Intuitively, the hybrid system requires several fine-tuning actions to adjust for user information needs accordingly. We then expected to observe differences in system performance and user satisfaction when users actively controlled the system, namely via keyword interactions (#$kw.int$ for short) and weighing of model parameters. Also, the only difference in control features between the two systems is the absence of model weight sliders in the baseline. Clearly, only those participants that manipulated the ranking weights in the hybrid tool were able to "appreciate" any difference with respect to the baseline (not consciously though due to between-subject setup).

With these considerations in mind, we established a clear cut to split participants of the hybrid system into 2 sub-groups: participants that performed at least one interaction with the ranking model, or "ranking interactions" for short, (Hyb_{WRI}, $N = 22$) and those that did not (Hyb_{NRI}, $N = 14$). Participants in the first group accumulated 5.05 ranking interactions on average ($sd = 5.19$, 17 of them with more than one). The two sub-groups represent manipulated variations, noted as objective system aspects (OSA) in the model diagram. Pathways originating from these nodes represent comparisons *against the baseline* ($N = 43$). We then built the structural model in Figure 6 based on the following hypotheses:

H1: The hybrid system will be significantly more accurate than the baseline (OSA). As we highlighted the importance of interactions for the system to understand user needs and thus be more accurate, we hypothesized higher system performance for the hybrid tool with ranking interactions, though influenced (at least to some extent) by the amount of keyword interactions ($Hyb_{WRI} \rightarrow$ #$kw.int \rightarrow aDCG$, where #$kw.int$ is an observed behavior or interaction (INT)).

H2: Participants will perceive higher degree of control and transparency with the hybrid system (SSA): The hybrid system provides features for enhanced control (model sliders in Figure 1.C) and transparency (tooltip and hint icons in Figure 3). At least users of the Hyb_{WRI} sub-group should perceive more control over the system and transparency provided by graphic explanations. Although at first we assumed that perceived result quality would mediate the effect of perceived control and transparency on satisfaction ($Perc.Control + Perc.Transp \rightarrow Perc.Res.Quality \rightarrow Satisfaction$), the fact that the latter had to be excluded from the model does not prevent us from measuring the direct effect ($Perc.Control + Perc.Transp \rightarrow Satisfaction$).

H3: Expertise will influence users' perception of the system (PC): Personal characteristics of a user are known to influence their experience with a system, but cannot be accounted for by the system itself [10, 24, 25]. As such, our assumption is that, regardless of the tool employed by a user, their level of familiarity with search interfaces will contribute to how controllable and transparent they perceive the system.

H4: The hybrid system will produce higher user satisfaction (EXP) Ultimately, we expected that objective (OSA) and subjective aspects (SSA) of the system would lead to higher user satisfaction (EXP). Although the direct link between system accuracy (OSA) and user experience (EXP) is often not evident or weak, Knijnenburg et al.'s framework [26] allows to model user experience as a result of how the user perceives the characteristics of the system ($OSA \rightarrow SSA \rightarrow EXP$).

4.2 Results

We applied structural equation modeling to analyze causal and mediated effects in behavioral (actions logs) and subjective data all together. The model in Figure 6 is the cornerstone to validate our hypotheses with respect to performance (Section 4.2.1) and user experience dimensions (Sections 4.2.2 and 4.2.4). The model obtained excellent goodness-of-fit, $\chi^2(141) = 146.55$, $p = .36$, $CFI = .998$, $TLI = .998$, $RMSEA = .022$ with $90\% CI = [0, 0.058]$. Regression

---

[3]In practice, some respecification of the measurement model is necessary. This is acceptable as long as decisions are not grounded in statistical considerations alone but rather in conjunction with theory and content considerations [5].
[4]using less than 3 indicators per factor is not advisable
[5]$AVE > .5$ indicates convergent validity. $\alpha > .8$ is good and $> .9$ is excellent
[6]Inter-factor correlations should be $< .85$
[7]χ^2 test should be non-significant. Acceptable cut-off values are $CFI > .96$, $TLI > .95$, $RMSEA < .05$ with upper-bound of $90\% CI < .10$

pathways report standardized coefficients (β) and robust standard errors (in parentheses).

4.2.1 System Performance. We found no difference in performance (*aDCG*) between the sub-group of hybrid system users with no ranking interactions and the baseline condition, $\beta = 0.06$, $SE = 0.11$, *ns* (denoted by the absence of links originating from the Hyb_{NRI} node). Conversely, when hybrid system participants interacted with ranking sliders, they performed significantly more keyword interactions (#*kw.int*), as it can be observed in the density distribution in Figure 7. In turn, more keyword interactions had a positive effect on system accuracy. Regression paths denote two types of effect of Hyb_{WRI} on *aDCG*: *(i)* a direct negative effect (path *c* in Figure 6), and *(ii)* an indirect positive effect through #*kw.int* (pathways *a* and *b*). Coefficients for indirect and direct effects with opposite signs indicate the possibility that the mediator (#*kw.int*) acts as a suppressor.

To measure the amount of mediation of Hyb_{WRI} on *aDCG* due to #*kw.int*, we first computed estimates for the indirect effect (product of *a* and *b*) and the total effect ($c + a * b$). Then we applied bootstrap[8] with 20,000 draws to obtain standard errors and 95% confidence intervals. We found a significant indirect effect, $\beta = 0.11$, $SE = 0.060$, $CI = [0.030, 0.277]$ (CI does not include 0, i.e. $p < .05$), while the total effect turned out non significant, $\beta = -0.16$, $SE = 0.106$, $CI = [-0.368, 0.051]$. A significant indirect effect that reduces the total effect to 0 indicates complete mediation. In other words, pathway *c* connecting Hyb_{WRI} with *aDCG* in Figure 6 is broken, and thus the entire effect of the independent variable (Hyb_{WRI}) on the dependent variable (*aDCG*) is transmitted through the mediator (#*kw.int*). This evidence supports **H1**.

Considering the caveat of careless interactions, we corroborated that #*kw.int* correlates with overall session time ($r = .5$, $p < .001$). Participants of the hybrid system with full interactions opened more document abstracts per bookmark (0.85 vs. 0.54), although not statistically significant ($U = 400$, $p = .08$). For visual assessment, Figure 8 plots the temporal progression of *aDCG* scores (with SE intervals) for the first 15 minutes of an average session of the 3 groups (baseline, Hyb_{NRI} and Hyb_{WRI}). Mean accuracy does not differ significantly, but it remains higher for Hyb_{WRI} throughout the average session ($M = 10'39''$, $SD = 5'53''$). To sum up, the evidence suggests that the hybrid system outperforms the baseline, *provided that users actively make use of control features*, i.e. interactions with keyword tags and ranking model sliders.

4.2.2 Perceived Control and Transparency. Participants did not perceive a significant difference in the *level of control* between the two systems, which could be attributed to the baseline having similar features to interactively control the document ranking (except for ranking sliders). In turn, the hybrid tool with ranking interactions (Hyb_{WRI}) had a direct positive effect on users' perception of *transparency*. Although both systems convey keyword contribution to overall scores (through stacked-colored bars), only the hybrid system provides explanations about the methods used to compute the document ranking. These results partially support **H2**. In other

---
[8]Bootstrap is a non-parametric technique based on resampling with replacement, often used for testing significance of indirect and total effects in mediation analysis [37].
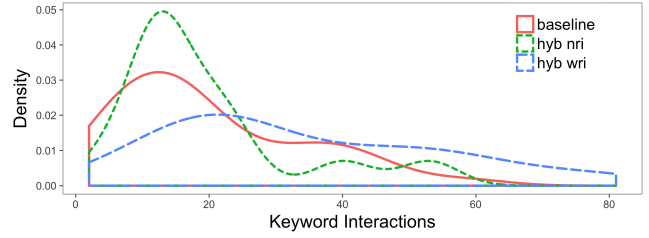

Figure 7: Density distribution of keyword interactions for the 3 groups analyzed in the user study. Approximately half of the participants of the hybrid SES with ranking interactions (*hyb wri*) performed over 40 keyword manipulations.
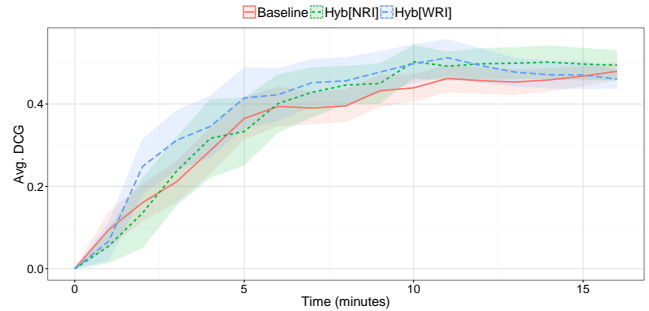

Figure 8: Temporal evolution of cumulative *aDCG*.

words, the SES system influenced the perception of transparency but not of control.

Moreover, we found no difference between the baseline and participants without ranking interactions (Hyb_{NRI}), which supports our initial assumption that the user should fully leverage the control features of the hybrid system in order to appreciate the benefit the combination of social and exploratory search.

4.2.3 Expertise. Regardless of the used system, users acknowledging high familiarity with search interfaces perceived a greater level of control. It is likely the case that experienced users understood these features better, as it has been observed that users perceive more control when they understand the system [24]. The model also reveals a large effect of perceived control on transparency. More specifically, perceiving a system as transparent seems to be greatly influenced by how controllable it is regarded by the user and indirectly by their search expertise, which provides evidence supporting **H3**.

4.2.4 User Satisfaction. Although sometimes the effect of system accuracy cannot be directly assessed in the experience (*EXP*), but through subjective perception of system aspects [10], in this case we found evidence of a direct effect. We modeled "satisfaction" as the final output of multiple objective and subjective variables. From the statistical evidence we can corroborate that higher satisfaction is a result of not only a good-performing system but also interaction and presentation aspects in the UI. Perceived control and transparency had each a positive effect on user satisfaction. Familiarity with search interfaces contributes to user satisfaction

because experienced participants felt more control over the system. This can be linked to previous findings reporting that domain experts are more satisfied when they can control the system [25]. It is important to highlight that the effect of the hybrid tool for the Hyb_{WRI} group on user satisfaction is not direct but transferable through perceived transparency (*SSA*) and system accuracy (*OSA*) (mediated by keyword interactions), thus supporting **H4**.

4.3 Discussion and Limitations

In this study we assessed system performance and user experience for a social exploratory search system contrasting it with a baseline system supporting exploratory search alone. Allowing the user to control not only keywords (a rather familiar action for any Web user), but also parameters of a hybrid ranking model, entails a higher level of complexity. Nevertheless, path and mediation analysis provided statistical evidence of better system performance for social exploratory search, as long as users exploit its full potential to refine both, keywords of interest and ranking parameters. Furthermore, higher system accuracy contributed directly to higher perceived transparency and higher user satisfaction.

Another interesting insight is that people who felt more in control also understood better the system's logic ($Perc.\,Control \rightarrow Perc.\,Transparency$). That is to say, the system is able to explain itself by way of interactions on top visual explanations. However, the model attributes a higher perception of control to users' personal traits rather than to the incremental customization in the hybrid system. It is worth noting that, despite higher complexity in the system, its was still as engaging as the exploratory search baseline. Users in SES condition that did not adjust the hybrid model (did not customize the fusion of content and social features), perceived similar levels of control and transparency, and were as satisfied as the users in the ES condition. This is indicated by the absence of pathways originating from Hyb_{NRI} in the structural model. In other words, the extra complexity was not detrimental to user experience. As using the control features gives a better understanding and appreciation of the SES, we would expect that by spending more time (beyond a single session), novice users will be able to assimilate and learn to exploit these advanced features. Alternatively, better initial support could speed up this process for novice users.

It is worth noting that the online nature of our study is at the same time an advantage and a limitation. On one hand, we believe the physical presence of an observer can lead to (intuitively positive) opinion bias. Moreover, through the crowd-sourced platform we managed to recruit more participants than we would have been able in a controlled in-lab studies. On the other hand, we lacked the opportunity to make our own observations and obtain deeper insights through post-study interviews.

Another aspect to be acknowledged is that the proposed model is one among infinite possibilities. This limitation is extrinsic and in general extends to any application of structural equation modeling. The hypotheses described in Section 4.1.6, although grounded in existing theories and common sense, reflect our own conceptualization of the user experience in a specific context. Thus, we do not claim that our model is absolute but a valid alternative.

5 CONCLUSIONS AND FUTURE WORK

In this paper we introduced an user-controlled approach for social exploratory search. Exploratory search was extended with social search capabilities by leveraging information left by past users, which in this case is bookmarking behavior. In turn, control and transparency, important features of many exploratory search systems, were used to integrate the social component in a way consistent with the exploratory search approach. On one hand, the user is empowered to shape their information needs and decide to what extent social information should be taken into account and, on the other hand, the system effectively communicates the effect of the user's decisions.

We conducted an empirical user study that reported evidence of the benefits of user-controllable fusion of content-based and social components of exploratory search (in the form of a hybrid interactive system). More specifically, the proposed structural model suggests that social exploratory search can indeed produce more accurate results, as long as the user actively "tells" the system how their information needs evolve and the right blend of content and social information. In turn, the explanatory nature of the UI had a positive effect on transparency. Together, system accuracy and perceived transparency promoted higher user satisfaction.

Conversely, the perception of control is mostly attributed to the users' own traits rather than differences among the two systems. However, despite the inherent increase in complexity introduced by the hybrid SES system, system performance and user experience were not affected for users that did not make use of the added functionalities. Based on this evidence, the challenge for our future work is to further exploit the interactive features to learn task models and improve ranking accuracy. Moreover, we plan to investigate how to deliver more suitable explanations of overall score composition for multiple sources of information.

REFERENCES

[1] J-w. Ahn and P. Brusilovsky. 2013. Adaptive visualization for exploratory information retrieval. *Inf. Process. Manag.* 49, 5 (2013), 1139–1164.

[2] J-w. Ahn, P. Brusilovsky, J. Grady, D. He, and R. Florian. 2010. Semantic Annotation Based Exploratory Search for Information Analysts. *Information Processing and Management* 46, 4 (2010), 383–402. DOI:http://dx.doi.org/10.1016/j.ipm.2010.02.001

[3] J-w. Ahn, P. Brusilovsky, D. He, J. Grady, and Q. Li. 2008. Personalized Web Exploration with Task Models. In *Proc. WWW '08*. ACM, 1–10.

[4] J-w. Ahn, R. Farzan, and P. Brusilovsky. 2006. Social search in the context of social navigation. *Journal of the Korean Society for Information Management* 23, 2 (2006), 147–165.

[5] J. C. Anderson and D. W. Gerbing. 1988. Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin* 103, 3 (5 1988), 411–423.

[6] R. A. Baeza-Yates, C. A. Hurtado, and M. Mendoza. 2004. Query Recommendation Using Query Logs in Search Engines. In *Current Trends in Database Technology - EDBT 2004 Workshops, Heraklion, Crete, Greece, March 14-18, 2004, Revised Selected Papers*. 588–596. DOI:http://dx.doi.org/10.1007/978-3-540-30192-9_58

[7] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. 2007. Optimizing web search using social annotations. In *the 16th international conference on World Wide Web, WWW '07*. ACM, 501–510. DOI:http://dx.doi.org/10.1145/1242572.1242640

[8] S. Bostandjiev, J. O'Donovan, and T. Höllerer. 2012. TasteWeights: A Visual Interactive Hybrid Recommender System. In *Proc. RecSys '12*. ACM, 35–42.

[9] R. Burke. 2002. Hybrid Recommender Systems: Survey and Experiments. *UMUAI* 12, 4 (Nov. 2002), 331–370.

[10] D. N. Chin. 2001. Empirical Evaluation of User Models and User-Adapted Systems. *UMUAI* 11, 1-2 (March 2001), 181–194.

[11] J. Chuang, C. D. Manning, and J. Heer. 2012. "Without the Clutter of Unimportant Words": Descriptive Keyphrases for Text Visualization. *ACM Trans. Comput.-Hum. Interact.* 19, 3, Article 19 (Oct. 2012), 29 pages. http://doi.acm.org/10.1145/2362364.2362367

[12] C. di Sciascio, V. Sabol, and E. E. Veas. 2015. uRank: Exploring Document Recommendations through an Interactive User-Driven Approach. In *Proc. RecSys '15 workshop IntRS*, Vol. 1438. 29–36.

[13] C. di Sciascio, V. Sabol, and E. E. Veas. 2016. Rank As You Go: User-Driven Exploration of Search Results. In *Proc. IUI '16*. ACM, 118–129. http://doi.acm.org/10.1145/2856767.2856797

[14] M. D. Ekstrand, D. Kluver, F. M. Harper, and J. A. Konstan. 2015. Letting Users Choose Recommender Algorithms : An Experimental Study. In *Proc. RecSys '15*. 11–18.

[15] B. M. Evans and E. H. Chi. 2010. An elaborated model of social search. *Information Processing and Management* 46, 6 (2010), 656–678. http://dl.acm.org/citation.cfm?id=1862565

[16] J. Freyne and B. Smyth. 2004. An experiment in social search. In *Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004) (Lecture Notes in Computer Science)*, Paul De Bra and Wolfgang Nejdl (Eds.), Vol. 3137. Springer-Verlag, 95–103.

[17] E. Gomez-Nieto, F. San Roman, P. Pagliosa, W. Casaca, E. S Helou, M. C. F de Oliveira, and L. G. Nonato. 2014. Similarity preserving snippet-based visualization of web search results. *IEEE TVCG* 20, 3 (March 2014), 457–70.

[18] C. He, D. Parra, and K. Verbert. 2016. Interactive Recommender Systems. *Expert Syst. Appl.* 56, C (Sept. 2016), 9–27.

[19] M. A. Hearst. 1995. TileBars: Visualization of Term Distribution Information in Full Text Information Access. In *Proc. CHI '95*. ACM Press, 59–66.

[20] M. A. Hearst. 2009. *Search User Interfaces* (1st ed.). Cambridge University Press.

[21] J. L. Herlocker, J. A. Konstan, and J. Riedl. 2000. Explaining collaborative filtering recommendations. In *Proc. CSCW '00*. 241–250. http://dl.acm.org/citation.cfm?id=358995

[22] O. Hoeber and X. D. Yang. 2008. Evaluating WordBars in exploratory Web search scenarios. *Information Processing and Management* 44, 2 (2008), 485–510.

[23] T. Joachims and F. Radlinski. 2007. Search Engines that Learn from Implicit Feedback. *Computer* 40, 8 (2007), 34–40. DOI:http://dx.doi.org/citeulike-article-id:3470248doi:10.1109/mc.2007.289

[24] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, and A. Kobsa. 2012. Inspectability and Control in Social Recommenders. In *Proc. RecSys '12*. ACM, 43–50.

[25] B. P. Knijnenburg, N. J. M. Reijmer, and M. C. Willemsen. 2011. Each to His Own: How Different Users Call for Different Interaction Methods in Recommender Systems. In *Proc. RecSys '11*. ACM, 141–148.

[26] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell. 2012. Explaining the user experience of recommender systems. *UMUAI* 22, 4-5 (2012), 441–504.

[27] G. Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41.

[28] A. Micarelli, F. Gasparetti, F. Sciarrone, and S. Gauch. 2007. *Personalized search on the World Wide Web*. Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York, 195–230.

[29] T. N. Nguyen and J. Zhang. 2006. A novel visualization model for web search results. *IEEE TVCG* 12, 5 (2006), 981–8.

[30] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams. 1993. Visualization of a document collection: The vibe system. *Inf. Process. Manag.* 29, 1 (1993), 69–81.

[31] D. Parra, P. Brusilovsky, and C. Trattner. 2014. See what you want to see: Visual User-Driven Approach for Hybrid Recommendation. In *Proc. IUI '14*. 235–240.

[32] J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. 2008. Algorithmic Mediation for Collaborative Exploratory Search. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08)*. ACM, New York, NY, USA, 315–322.

[33] P. Pirolli and S. Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. International Conference on Intelligence Analysis*.

[34] T. Ruotsalo, J. Peltonen, M. Eugster, D. Glowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki, G. Jacucci, and S. Kaski. 2013. Directing exploratory search with interactive intent modeling. In *Proc. CIKM '13*. 1759—-1764.

[35] G. Shani and N. Tractinsky. 2013. Displaying relevance scores for search results. *ACM SIGIR '13* (2013), 901–904.

[36] B. Shneiderman, D. Byrd, and W. B. Croft. 1998. Sorting out Searching: A User-interface Framework for Text Searches. *Commun. ACM* 41, 4 (April 1998), 95–98.

[37] P. E. Shrout and N. Bolger. 2002. Mediation in experimental and nonexperimental studies: New procedures and recommendations. (2002).

[38] R. W. White, B. Kules, S. M. Drucker, and others. 2006. Supporting exploratory search. *Commun. ACM* 49, 4 (2006), 36–39.

[39] M. L. Wilson, b. Kules, M. C. Schraefel, and B. Shneiderman. 2010. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science* 2, 1 (2010), 1–97.

[40] K-P. Yee, K. Swearingen, K. Li, and M. A. Hearst. 2003. Faceted metadata for image search and browsing. In *ACM Conference on Human Factors in Computing Systems, CHI 2003*. ACM Press, 401–408. http://citeseer.ist.psu.edu/yee03faceted.html

[41] Z. Yue, S. Han, and D. He. 2014. Modeling Search Processes Using Hidden States in Collaborative Exploratory Web Search. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 820–830. DOI:http://dx.doi.org/10.1145/2531602.2531658