# PaperExplorer: Personalized Exploratory Search for **Conference Proceedings**

Behnam Rahdari, Peter Brusilovsky

School of Computing and Information, University of Pittsburgh, 135 North Bellefield Avenue Pittsburgh, PA 15260, USA

#### Abstract

This paper presents our attempt to create an exploratory search system, PaperExplorer, for a historic archive of conference proceedings. PaperExplorer uses concept extraction, knowledge graphs, and user-controlled recommendation to assist users with various levels of domain expertise in their information needs.

#### Keywords

Exploratory Search, Knowledge Graph, Information Exploration, Intelligent interface

## 1. Introduction and Background

Exploratory search systems form an increasingly popular category of information access and exploration tools. These systems creatively combined search, browsing, and information analysis steps shifting user efforts from recall (formulating a query) to recognition (i.e., selecting a link) and helping them to gradually learn more about the explored domain [1].

In this paper we present our attempt to augment the set of search systems focused on conference proceedings with a personalized exploratory search system PaperExplorer <sup>1</sup>. We hope that PaperExplorer ability to support information discovery, learning-while-searching, and personalization could help a broader set of users to benefit from the assembled collection of conference proceedings.

## 1.1. Exploratory Search

A number of real-life search tasks require a considerable amount of learning during the search process to achieve adequate results. These tasks are known as *exploratory* search tasks [2]. Since simple search systems are usually not efficient in supporting exploratory search tasks, a range of specialized systems have been developed and evaluated.

More recently, few projects in this area demonstrated that the effectiveness of exploratory search could be improved by using a personalized system, which builds a profile of user interests and adapts to the individual user [3]. The work presented in this paper investigates the ideas of profile-based exploratory search in the context of finding research publications related to a certain conference.

#### 1.2. Controllability

User controllability has been recognized as a valuable component of advanced information access interfaces. The ideas of controllability were made popular by a stream of work on user-controllable recommender systems [4]. However the value of extended user control has been also demonstrated in the area of exploratory search.

For example, NameSieve [5] presented a summary of search results in the form of entity clouds, which a controllable filtering and exploration of results. PeopleExplorer [6] offered users an option to re-sort people search results based on multiple user-related factors. uRank [7] introduced a controllable interface for refining and reorganizing search result and SciNoon [8] simplifies the exploratory search process for scientific groups.

## 1.3. Open User Profile

The idea to apply open user profiles (also known as open user models) to better support personalized information access was among the early ideas explored in this field. Open user profiles allow users to examine and possibly change the content of their interest profiles, which are used to personalize their search or browsing process.

Since the open user profiles increase interactivity, transparency, and controllability of the information exploration process, their application was a good match to the nature of exploratory search. While first attempts to introduce "bag-of-words" open user profiles had mixed success [9], more recent work focused on semantic level user profiles demonstrated its potential for personalized exploratory search [3, 10].

We start the paper with the presentation of PaperExplorer interface and follow with the details on concept

Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station USA

<sup>🛆</sup> ber58@pitt.edu (B. Rahdari); peterb@pitt.edu (P. Brusilovsky) D 0000-0001-6514-912X (B. Rahdari); 0000-0002-1902-1464

<sup>(</sup>P. Brusilovsky) © Constraints (CEUR Workshop Proceedings (CEUR-WS.org) (CEUR Workshop Proceedings (CEUR-WS.org)

<sup>&</sup>lt;sup>1</sup>http://scythian.exp.sis.pitt.edu/ht/

A	Q, Start you	Start your search here												
в	Similar Keywords: social networks		Pearson correlation		on	? + social media		<b>?</b> + 9		ic algorithm	?+	social media pl	atforms	?+
P Hint : Use  to add a keyphrase directly to your profile of interests. Use  to get more information about a keyphrase from Wikipedia. Use  to add a keyphrase for removal. Use the Siders to tune the importance of keyphrases														
C	recomm	nender systems	collabor	ative filtering	× algori E	ithm >								
s	ack	Title										Details	Туре	Year
		+ a collaborative filtering recommendation based on genetic algorithm and self-organized map for learning material									More Details	paper	2009	
D	01	- evaluation of calaborative fibering algorithms for recommending articles on citacular      - Abstract:     - modivated by the potential use of calaborative tagging systems to develop new inscommending systems, we have implemented and compared three variants of user-based calaborative fibering algorithms to     provide recommendations of an assurance classic calaborative tagging systems to     an improvement of precision, and     an improvement of precision, and     an expression correlation to calculate the similarly between users and there registronic users     and advectory tagging     an advectory tagging     an advectory tagging systems     an advectory tagging system     an expression of the size of precision, and     an expression tagging system     and     an advectory tagging     and     an advectory tagging     and     an advectory tagging     and     and									More Details	paper	2009	
	social networks and interest similarity: the case of citeulike											More Details	short paper	2010
matrix algorithms in social recommendation systems												More Details	tutorial	2014

Figure 1: Interface Design of Paper-Explorer representing different parts of the system.

extraction, knowledge graph organization, and recommendation that enable the work of this interface.

# 2. The Interface of PaperExplorer

Personalized information exploration in PaperExplorer is centered around user interest profile [11] - a collection of concepts represented by *keyphrases* that express user interests. Unlike traditional search that requires users to specify all keyphrases in a query, PaperExplorer supports users in the process of gradual discovery and refinement of their interests. It also allows the users to control the importance of each keyphrase in recommending relevant results. PaperExplorer interface consists of the following main sections.

## 2.1. Instant Search Box

The search box (Figure 1A) is the gateway to the system. The instant search approach allows users to discover relevant keyphrases representing concepts of interest without a fully formulated query. When a user starts typing a query, a series of matching keyphrases appears helping the user to discover a concepts of interest (e.g., User Interfaces and User Modeling). When an item is selected from the list, it will automatically adds to the slider area (Figure 1C). at the same time, an updated list of search results will be presented to the user.

#### 2.2. Recommended Keyphrases

When at least one keyphrase is added to the user's profile, the system recommends five semantically similar concepts (shown as keyphrases) in the *Similar keyphrases* area of the interface (Figure 1B). Users can add recommended keyphrases to their interest profiles by clicking on the plus button to the right of each keyphrase. As the user's profile grows and refines, the set of recommended concepts is updated since the system recommends instances similar to all concepts in the user's profile. Each recommended concept also provides users with a short description of the concept. Clicking on the question mark button next to the add button, opens up a separate window containing the abstract of that concept's Wikipedia entry.

#### 2.3. Open User Profile

The slider area (Figure 1C) displays the current user profile of interest. PaperExplorer implements a contentbased recommendation approach, which generates the list of recommended results (Figure 1D) using the profile. To support transparency and controllability of this process, the interest profile is visible and directly editable by the end users.

To build the profile the user can add relevant concepts represented by keyphrases as explained above as well as remove less relevant keyphrases (using the red x) as they discover more relevant concepts or explore different interests.

Sliders associated with each keyphrase enable users to control the relative importance of the represented concept compared to others in their profile, ranging from 1 (least important) to 10 (most important). The use of sliders for fine-tuning of user profile was motivated by keyword tuning approach in uRank [7], which was confirmed as a user-friendly and efficient in an exploratory search context. All actions within the profile (adding, removing, or adjusting sliders) immediately affect the search results list.

## 2.4. Search Results

As soon as the user adds the first keyphrase to the interest profile, a table of the 20 most relevant publications



Figure 2: Graph Schema representing the entities of the knowledge graph and the relationship between them

is generated (Figure 1:D). The first column of the table visualizes the combined relevance between keyphrases in the user interest profile and each result. The colors in the stacked-bar (Figure 1:D1) are matched with the color of slider in the profile and the size and opacity of each bar expresses the relevance of the result to each profile keyphrase.

The second column of table lists the titles of relevant publications. Clicking on each title expands a window that holds the abstract of the paper. The mentioned keyphrases are highlighted with corresponding colors.

The opacity of the colors reflect the relevance of a keyphrase to the paper and the current value of slider for that keyphrase. To further assist the users, PaperExplorer underlines all available keyphrases in the text (both in title and abstract).

Hovering over the underlined portion of the text opens a popup window (Figure 1:D2) that enable user to (1) see the relevance of the keyphrase to the text in a form of a vertical bar-chart, (2) add the keyphrase directly to the interest profile, and (3) report the improper keyphrases to the administrator for removal.

The latter helps us to improve the quality of extracted keyphrases and eliminate the occasional errors in the process of extraction.

## 3. The Knowledge Graph

The knowledge graph consists of three main entities publications, authors, keyphrases and their relationships - extracted from our data set and hosted in a native graph database Neo4j<sup>2</sup>.

Figure 2 presents the schematic representation of the knowledge graph. Authors are interconnected by the relation *Co-Author* (based on co-authorship) and connected to papers by the relation *Published*. Papers connected to keyphrases using the *Has-Key* relationship. The latter carries a weight that determines the strength of the relationship between each keyphrase and the publication.

## 3.1. Data Source and Keyphrase Extraction

We used the collection of proceedings from two main conferences (Hypertext and UMAP) as the main source of data to build the knowledge graph and extract the keyphrases. This collection covers all publications of these two conferences from 2008 to 2020. Using this dataset and the concept extraction explained below, we generated the knowledge graph covering 2023 publications. 14404 keyphrases were extracted from titles and abstracts of these publications.

We used TopicRank [12], a graph-based keyphrase extraction method to extract the initial set of candidate keyphrases from the title and abstract of the publications. We then used the Wikipedia API to filter all extracted keyphrases; only keyphrases with an entry in Wikipedia were kept in the knowledge graph. We further assign weight to each publication keyphrase pair using cosine similarity between the bags-of-words extracted from the Wikipedia page and the publications.

## 4. Profile-Based Search

We deployed a two-phase search process to produce the most relevant results based on user interest profile. In the first phase, a primary list of candidates is being selected from the graph and the second phase assure that the results are presented to the user in the right order based on their relevancy to the query. We describe these two phases in more details in the following.

*Candidate selection*: We used the Cypher Querying Language to generate the initial list of candidate publications. At each instance of user interaction with the system (e.g., adding/removing keyphrases or tuning the sliders), the system considers all publications connected to at least one of the concepts of interest in the user profile.

*Reordering the results*: After generating the list of candidate results, the system rearranges the results in a way that the most relevant results appear at the top of the list. In order to do that, first a complete list of keyphrases that appear in the text (title and abstract) of each publication, alongside with their relevancy score (weight) is being generated. Then for every keyphrase that exist in the user interest profile, we multiplied its weight with the value of corresponding slider. Finally, the relevance score is assigned to each candidate considering candidate's similarity to each of profile concepts and the value of the sliders.

<sup>&</sup>lt;sup>2</sup>https://en.wikipedia.org/wiki/Neo4j

## 5. Experience and Future Work

PaperExplorer system has been deployed online and also demonstrated to several target users. The early results indicate that the success of the system to a considerable extent depends on the quality of keyphrase extraction. We are interested to collaborate with experts on keyphrase extraction to develop approaches optimized for exploratory search.

## References

- R. W. White, B. Kules, S. M. Drucker, et al., Supporting exploratory search, Communications of the ACM 49 (2006) 36–39.
- [2] G. Marchionini, Exploratory search: From finding to understanding, Communications of the ACM 49 (2006) 41–46.
- [3] F. Bakalov, B. König-Ries, A. Nauerz, M. Welsch, IntrospectiveViews: An interface for scrutinizing semantic user models, in: 18th International Conference on User Modeling, Adaptation, and Personalization, Springer, 2010, pp. 219–230.
- [4] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, A. Kobsa, Inspectability and control in social recommenders, in: 6th ACM Conference on Recommender Systems, 2012, pp. 43–50.
- [5] J.-w. Ahn, P. Brusilovsky, J. Grady, D. He, R. Florian, Semantic annotation based exploratory search for information analysts, Information Processing & Management 46 (2010) 383–402.
- [6] S. Han, D. He, J. Jiang, Z. Yue, Supporting exploratory people search: a study of factor transparency and user control, in: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, ACM, 2013, pp. 449–458.
- [7] C. di Sciascio, V. Sabol, E. E. Veas, Rank as you go: User-driven exploration of search results, in: 21st International Conference on Intelligent User Interfaces, 2016, pp. 118–129.
- [8] Y. Nedumov, A. Babichev, I. Mashonsky, N. Semina, Scinoon: Exploratory search system for scientific groups, in: IUI 2019 Workshop on Exploratory Search and Interactive Data Analytics, 2019. URL: http://ceur-ws.org/Vol-2327/ IUI19WS-ESIDA-3.pdf.
- [9] J.-w. Ahn, P. Brusilovsky, J. Grady, D. He, S. Y. Syn, Open user profiles for adaptive news systems: help or harm?, in: the 16th international conference on World Wide Web, WWW '07, ACM, 2007, pp. 11–20.
- [10] T. Ruotsalo, G. Jacucci, S. Kaski, Interactive faceted query suggestion for exploratory search: Whole-

session effectiveness and interaction engagement, Journal of the Association for Information Science and Technology 71 (2020) 742–756.

- [11] B. Rahdari, P. Brusilovsky, D. Babichenko, Personalizing information exploration with an open user model, in: 31st ACM Conference on Hypertext and Social Media (HT '20), Association for Computing Machinery, New York, NY, USA, 2020, p. 0. doi:10.1145/3372923.3404797.
- [12] A. Bougouin, F. Boudin, B. Daille, TopicRank: Graph-based topic ranking for keyphrase extraction, in: Proceedings of the Sixth International Joint Conference on Natural Language Processing, Asian Federation of Natural Language Processing, Nagoya, Japan, 2013.