

Telling How to Narrow it Down: Browsing Path Recommendation for Exploratory Search

Mostafa Dehghani¹
Alex Olieman¹

Glorianna Jagfeld²
Jaap Kamps¹

Hosein Azarbonyad³
Maarten Marx³

¹Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

²Institute for Natural Language Processing, University of Stuttgart, Germany

³Informatics Institute, University of Amsterdam, The Netherlands

{dehghani,h.azarbonyad,olieman,kamps,maartenmarx}@uva.nl, glorianna.jagfeld@ims.uni-stuttgart.de

ABSTRACT

Supporting exploratory search tasks with the help of structured data is an effective way to go beyond keyword search, as it provides an overview of the data, enables users to zoom in on their intent, and provides assistance during their navigation trails. However, finding a good starting point for a search episode in the given structure can still pose a considerable challenge, as users tend to be unfamiliar with exact, complex hierarchical structure. Thus, providing look-ahead clues can be of great help and allow users to make better decisions on their search trajectory.

In this paper, we investigate the behavior of users when a recommendation engine is employed along with the browsing tool in an exploratory search system. We make use of an exploratory search system that facilitates browsing by mapping the data on a hierarchical structure. We designed and developed a path recommendation engine as a feature for this system, which given a text query, ranks different browsing paths in the hierarchy based on their likelihood of covering relevant documents. We conduct a user study comparing the baseline system with the featured system.

Our main findings are as follows: We observe that, using the baseline system the users tend to explore the data in a breadth-first-like approach by visiting different data points at the same level of abstraction to choose one of them to expand and go deeper. Conversely, with browsing path recommendation (BPR) as a feature, the users tend to drive their search in a more depth-first-like approach by quickly going deep into the data hierarchy. While the users still incline to explore different parts of the search space by using BPR, they are able to restrain or augment their search focus more quickly and access smaller but more promising regions of the data. Therefore, they can complete their tasks with less time and effort.

Keywords: Exploratory Search, Recommendation, BPR, ExPoSe.

1. INTRODUCTION

There are several information needs requiring sophisticated human-computer interactions that currently remain unsolved or poorly

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR '17, March 07 - 11, 2017, Oslo, Norway

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4677-1/17/03...\$15.00

DOI: <http://dx.doi.org/10.1145/3020165.3022155>

supported by major search applications. One of these cases is exploratory search, which refers to search tasks that are open-ended, multi-faceted, and iterative, like learning or topic investigation [5]. This type of search often occurs in a domain unknown or poorly known to the searchers which can make it hard for them to formulate proper queries for retrieving useful documents [3].

Exploratory search is composed of two main activities, *exploratory browsing* and *focused searching* [9]. Exploratory browsing refers to activities that aim at better defining the information need and raising the understanding of the information space. Focused searching corresponds to activities like query refining and results' comparisons after the information need has been shaped more clearly. Based on this composition, an exploratory search system needs to provide its users a connected space of information to browse and investigate, as well as facilities to adjust the focus of their search towards useful documents.

Using structured data to organize unstructured information is one of the promising approaches for supporting complex search tasks, including exploratory search [6]. Structure in the data provides overviews at different levels of abstraction and empowers the users to explore the data from different points of view. However, it may still be difficult to find useful paths of exploration and clues can help the users [1].

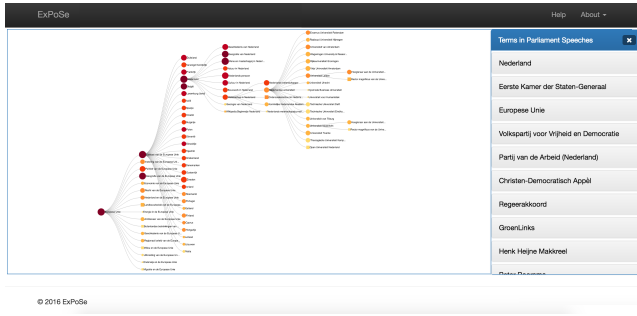
The main aim of this paper is *to investigate the user behavior in exploratory search when a recommendation engine for browsing paths is provided along with the browsing system*. To do so, we have employed the *ExPoSe-Browser* (Exploratory Political Search Browser) as the baseline system and built a recommendation engine as a supplementary feature for the system. We have conducted a user study involving exploratory search tasks which revealed general differences of the browsing behavior of the subjects using the two different systems.

We break down the main goal of the paper into two concrete research questions:

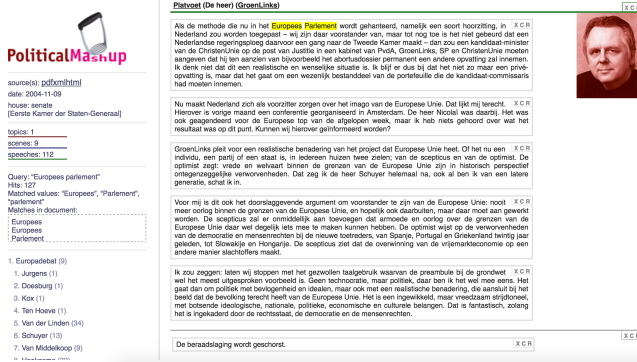
RQ1 *What is the effect of providing browsing path recommendations on user browsing behavior in exploratory search?*

RQ2 *Does browsing path recommendation assist users in exploratory search?*

We demonstrate that in general, providing suggestions for paths for browsing and exploring the data helps the users to not only discover more promising regions of the search space and be more satisfied, but also to quickly narrow down their search, leading to shorter search sessions. In the rest of this paper, we first introduce the case study system in Section 2. Then we describe our experimental setups and presents the results in Section 3. In Section 4, we conclude.



(a) A sample expanded hierarchy for exploring Dutch parliamentary proceedings in search of “European union”. Left panel provides the hierarchies of topics based on Wikipedia categorical structure and right panel demonstrates the entities and speeches related to the selected topic.



(b) A sample of visited speech by a user in search of “European union”. The page contains information and the content of whole debate and the searched entity is highlighted.

Figure 1: Exploratory Political Search (ExPoSe) System.

2. BPR

Browsing the search space is an important element in exploratory search. While organizing data into a structured format improves the browsability, difficulties on deciding the direction of exploration might remain. In this section, we introduce the ExPoSe-Browser as the system we employed in our case study and explain the BPR engine we added as a supplementary feature to the system.

2.1 ExPoSe-Browser System

The ExPoSe-Browser [2, 4] is an exploratory search system that provides a hierarchical grouping of parliamentary speeches based on Wikipedia categories. More precisely, it enables its users to browse and investigate the parliamentary proceedings from an arbitrary point of view. First, the user has to choose the point of view from the set of Wikipedia category. Then, the ExPoSe-Browser makes a projection of the parliamentary speeches to the Wikipedia categorical structure with the point of view as the root of the hierarchy. This empowers the user to dive into the data from an abstract level to the detailed information by traversing the hierarchical structure.

The projection process is based on Wikipedia named entities mentioned in the parliamentary speeches. A customized entity linker for parliamentary conversation [7, 8] is employed to recognize and disambiguate the general named entities as well as domain specific entity types like politicians and political parties from all the speeches. Then ExPoSe-Browser lets the user select a sub-hierarchy from Wikipedia’s category hierarchy¹ onto which mentioned entities

¹Formally, Wikipedia’s category system is modeled as a directed acyclic graph rather than a tree structure.

from the text are projected. In the selected hierarchy, each node is a category which represents a topic. Its descendant nodes are its sub-topics. Parliamentary speeches are grouped into the nodes in the hierarchy based on the categorical memberships of entities mention in them.

Based on the debates and entities related to the categories in the hierarchy, the system calculates the importance and recency of each category. The importance of each node demonstrates how much the topic of that category is addressed in the parliamentary debates, based on the frequency of entities related to this category. The recency of each node shows how recently the topic of this category was discussed in the debates. Importance and recency are visualized by the size and color of the nodes in the user interface. Nodes can be expanded to display the next level of the hierarchy. Furthermore, the list of the speeches categorized under each node can be accessed as well as the full content of the speeches with the related entities highlighted in the text. In summary, the ExPoSe-Browser provides an overview of the data at different levels of abstraction and implicitly offers a faceted interface by clustering the data into different topical aspects. Moreover, the users can explore multiple alternative browsing paths by going-back-and-forth functionalities.

As an example of the type of complex search that can be addressed with ExPoSe-Browser, assume a user is interested in analyzing the relation between national laws of the Netherlands and European union legislation. A reasonable approach is to investigate when EU legislation was discussed in the Dutch parliament and in the context of which (proposed) Dutch laws. In other words, it is desirable to see how debates within the parliament of the Netherlands can be projected onto the topics related to the European Parliament in terms of both the subject matter and time. Using the ExPoSe-Browser, user is able to browse and investigate Dutch parliamentary proceedings regarding topics related to EU from a very abstract level (categories - Figure 1a) to a very detailed level (entities/content of speeches - Figure 1b).

2.2 Path Recommendation

We design and implement a recommendation engine for suggesting paths of browsing and exploration as a supplementary plugin for the ExPoSe-Browser system. In general, the BPR engine, given a text query, provides a ranking of paths based on their likelihood of leading to speeches relevant to the query (Figure 1).

More precisely, the recommendation engine, for a given text query, calculates the similarity of the query with the full content of all speeches. Then, it assigns a score to each entity based on its occurrence in different speeches and their corresponding scores. Afterwards, it integrates the normalized scores of all entities under each category as the likelihood score of the category given the searched query. Finally, the likelihood score is mapped to the weight of the path from the root of the hierarchy to the category and the thickness of each edge in the hierarchy is determined by the weights of all paths that pass through the edge.

Providing a ranking of paths to speeches instead of ranking the speeches themselves can be considered as a preview clue that brings a limited amount of certainty to the exploration process while its bias is not strong enough to completely deter the users from exploring the space.

3. EXPERIMENTS

In this section, we present a user study we conducted to investigate the effect of BPR on user behavior in exploratory search. First, we describe our experimental setups, then we present the results of the experiments.

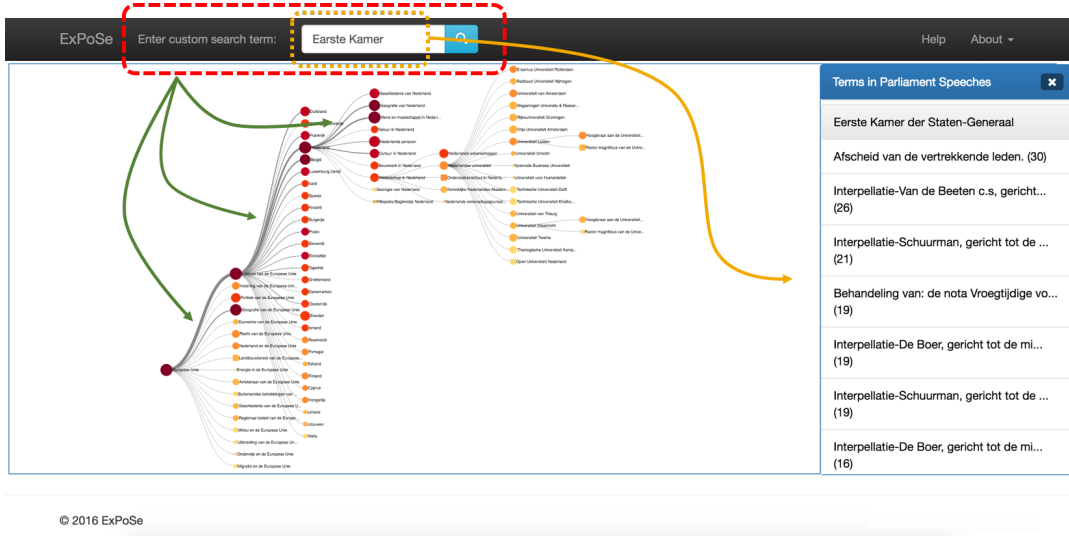


Figure 2: BPR for Exploratory Political Search Browser (ExPoSe-Browser) System.

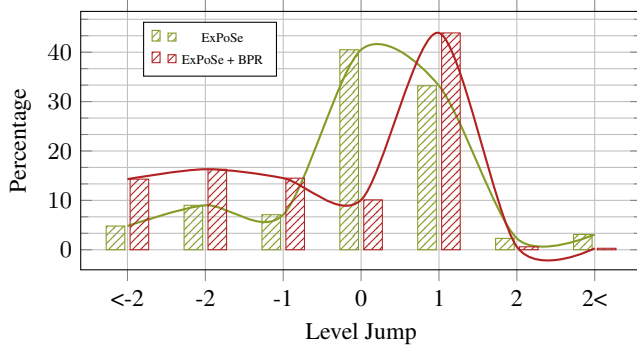


Figure 3: Distribution of different level jumps: positive and negative jumps mean toward lower and higher levels respectively.

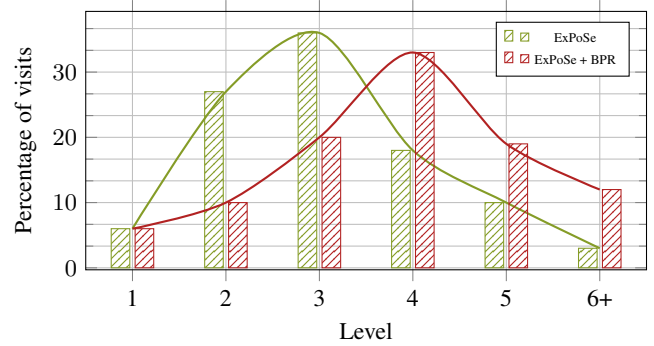


Figure 4: Distribution of level visits.

3.1 Experimental Setup

Twelve participants were enrolled in this study who had an average age of 27 (SD = 2.14). Four participants were male and eight female. Three participants were undergraduate students, nine were graduate students. All participants had more than 10 years of computer experience and a fairly high ability of online search. All participants were inexperienced in the political domain. We compiled eight general search tasks for which there is no concrete answer. Briefly, the tasks were to summarize the attitude/opinion of the Dutch parliament on specific topics: *Immigration, Islam, World War II, Tax, Holocaust, Dutch Golden Age, EU Membership of the UK, and EU membership of Greece*.

Each participant was given two different tasks to complete using the baseline system and then two other tasks to complete using the featured system. The tasks were distributed among the participants and systems in such a way that each task was completed three times with the baseline system and three times with the featured system, each time by different participants. The overall number of search sessions was 48, consisting of an equal number of search sessions with the baseline and featured system. Prior to working on the tasks, the users were trained to work with the systems and got to know their features. During the experiments, they were given a short and clear description of the tasks. They were asked to provide a summary of two to three paragraphs of the information obtained in

each search session. All interactions of the users with the systems were logged during the experiments.

3.2 Effect of Path Recommendation

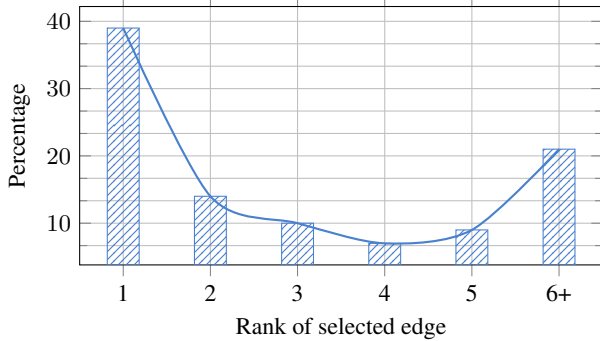
In this section, first, we address our first research question: “What is the effect of providing browsing path recommendations on user browsing behavior in exploratory search?”

In each search session, the users could to move back-and-forth across the hierarchy and visit different nodes at different levels. In order to investigate and compare the node visits in two systems, for each system, we count the number of different level jumps across all sessions. The level jump can be zero, which means the currently visited node is at the same level as the last visited node. A positive level jump indicates that the currently visited node is at a lower level than the last visited node (forward move). A negative level jump means the currently visited node is at a higher level than the last visited node (backward move).

Figure 3 shows the percentage of different level jumps in both systems. According to the plots in the figure, in the baseline system, users tend to explore the nodes at one level (high percentage of zero-jumps) and then choose one of them to go to the next level, which is a breadth-first-like traversal model of the data. However, in the featured system with BPR, the users mostly choose to quickly descend in the hierarchy, which is a more depth-first-like traversal model of the data. Moreover, in the featured system, there is a fairly high percentage of negative jumps compared to the baseline system,

Table 1: Statistics from search sessions for both systems. The given numbers are per-session statistics, averaged over all search sessions.

System	avg. num. of root selection	avg. num. of node expansion	avg. num. of entity list loading	avg. num. of speech list loading	avg. num. of speech content loading	avg. session duration (min)
ExPoSe	1.2 (SD=0.33)	16.4 (SD=4.1)	29.5 (SD=8.4)	16.7 (SD=8.7)	36.3 (SD=8.7)	31.6 (SD=7.5)
ExPoSe + BPR	1.3 (SD=38)	9.7 (SD=2.2)	21.1 (SD=3.16)	12.4 (SD=2.5)	18.4 (SD=3.7)	23.1 (SD=6.1)

**Figure 5:** Distribution of ranks of selected edges.

which demonstrates that users backtrack to try other paths from higher levels, probably to cover different aspects of the topic.

Figure 4 presents the distribution of node visits over different levels during the search sessions for both systems. Based on the plots, in average, users visited deeper levels of the hierarchies when they used the featured system. Figures 3 and 4 demonstrate that BPR makes users more confident to quickly narrow down their search space and deepen their exploration path. Additionally, they feel less risk to conduct back-and-forth moves in the search space.

In another analysis, we investigate the chance of an edge being selected to be pursued by the user according to its rank given from the BPR. Figure 3 demonstrates the percentage of times that edges in different ranks are selected. As can be seen, although edges with the top rank were selected in almost 40% of cases, the users also tried paths with lower ranks. Thus, BPR did not decrease the chance of exploration.

Now, we address our second research question: “Does browsing path recommendation assist users in exploratory search?”

We extracted statistics from all search sessions to study the difference between two systems in terms of the search system quality and user satisfaction. Table 1 presents the statistics of different actions per session by users averaged over all search sessions in both systems. As can be seen, in the featured system, the average number of nodes expanded in the hierarchy in each session is much lower compared to the baseline system. Moreover, the average number of requests for displaying more detailed information like the list of entities, the list of speeches, or content of speeches under each node is lower than the baseline system. The average duration of search session is also lower in the featured system.

Furthermore, we asked two experts in the political domain to rate the quality of the summaries generated by the users as an indicator for their understanding of the topic on a 10-point scale. Summaries generated using the featured system get fairly higher ratings ($M=7.73$, $SD=2.10$), compared to the summaries generated using the baseline system ($M=5.36$, $SD=3.31$).

Considering the statistics from Table 1 along with the ratings given to the summaries generated by the users, we conclude that with BPR, users can complete their complex tasks better and by making less effort and spending less time.

4. CONCLUSIONS

The main aim of this paper was to investigate the user behavior

in exploratory search when a recommendation engine for browsing paths is provided along with the browsing system.

We designed a BPR engine and employed it in the ExPoSe-Browser system, our baseline exploratory search system to investigate the change in the behavior of the users. We observed that providing BPR leads to a depth-first-like traversal by users since it brings some certainty into the decision-making process of the users and makes them able to quickly deepen their search. However, without path recommendation, they need to explore different options at each level which corresponds to a breadth-first-like strategy for exploration. We also found that using BPR, users still tend to explore different parts of the search space. But they adjust their search focus to make less effort and spend less time for completing their tasks.

In future work, we can improve the BPR engine by making it adaptive to the user’s browsing history. As the analytical future work, we are going to study the effect of the complexity of tasks in the same setting. We are also interested to study to which extend providing recommendations in the different parts of an exploratory search system like query suggestion, node expansion hints, etc. helps.

Acknowledgments

This research is funded in part by Netherlands Organization for Scientific Research through the *Exploratory Political Search* project (ExPoSe, NWO CI # 314.99.108), and by the Digging into Data Challenge through the *Digging Into Linked Parliamentary Data* project (DiLiPaD, NWO Digging into Data # 600.006.014).

References

- [1] A. Azman and I. Ounis. Browsing recommendation based on the intertemporal choice model. In *International Conference on Flexible Query Answering Systems*, pages 215–226, 2009.
- [2] M. Dehghani, A. Olieman, and H. Azarbonyad. Discovering links in political conversations. Technical report, University of Amsterdam, 2015.
- [3] A. Diriye, M. L. Wilson, A. Blandford, and A. Tombros. Revisiting exploratory search from the hci perspective. In *HCIR 2010*, pages 99–102, 2010.
- [4] ExPoSe. Exploratory political search project. <http://humanities.uva.nl/~kamps/expose/>, 2015. Netherlands Organization for Scientific Research.
- [5] G. Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, 2006.
- [6] N. Marie. *Linked data based exploratory search*. PhD thesis, Université Nice Sophia Antipolis, 2014.
- [7] A. Olieman, H. Azarbonyad, M. Dehghani, J. Kamps, and M. Marx. Entity linking by focusing dbpedia candidate entities. In *Proceedings of the First International Workshop on Entity Recognition & Disambiguation, ERD’14*, pages 13–24, 2014.
- [8] A. Olieman, J. Kamps, M. Marx, and A. Nusselder. A Hybrid Approach to Domain-Specific Entity Linking. In *Joint Proceedings of the Posters and Demos Track of 11th International Conference on Semantic Systems (SEMANTICS’15) and 1st Workshop on Data Science: Methods, Technology and Applications (DSci’15)*, pages 55–58, Vienna, 2015. URL <http://arxiv.org/abs/1509.01865>.
- [9] R. W. White and R. A. Roth. Exploratory search: beyond the query-response paradigm (synthesis lectures on information concepts, retrieval & services). *Morgan and Claypool Publishers*, 3, 2009.