Evaluation of Personalized Concept-Based Search and Ranked Lists over Linked Open Data

Melike Sah and Vincent Wade
Centre for Global Intelligent Content, Trinity College Dublin, Dublin, Ireland
{Melike.Sah, Vincent.Wade}@scss.tcd.ie

Abstract. Linked Open Data (LOD) provides a rich structured data. As the size of LOD grows, accessing the right information becomes more challenging. Especially, the commonly used ranked lists presentation of current LOD search engines is not effective for search tasks in unfamiliar domains. Recently, combination of clustering and personalized search gained more attention for this purpose. In this paper, we evaluate the impact of personalized concept-based search in terms of task assistance and user satisfaction, while comparing with the ranked lists. A user study was conducted with 32 subjects. Results showed that the personalized concept-based search enabled users to be more effective and efficient at performing both information gathering and fact finding tasks.

Keywords: Click through, personalized search, linked open data, ranked list.

1 Introduction

As the size of LOD grows, providing efficient search mechanisms becomes important. However, current LOD search engines (e.g. Sindice [1], Watson [2]) use ranked lists presentation, which is not efficient for information gathering search tasks in unfamiliar domains [3]. This is an important drawback since it is estimated that ~80% of Web queries are informational [11]. Faceted search overcomes this issue by assisting users with topics for interactive search and browsing [4]. However, facet creation depends on specific data and schema properties of underlying metadata and it can be difficult to generate useful facets to large and heterogeneous LOD [5]. In traditional Information Retrieval (IR), results clustering and personalized search are two popular methods for enhancing search efficacy. In clustering search, results are organized into categories for assisting users in results exploration and in query disambiguation. Results categorization is used widely such as Google categories or Yahoo Directories. Alternatively, personalized search aims to enhance retrieval by adapting results to context/interests of the user. On the other hand, [6] and [7] combine both results clustering and personalization in order to improve retrieval accuracy. [6] uses hierarchical clustering and user's manual selections on these clusters to filter out relevant results. Whereas, [7] assumes that search results are clustered (such as using ODP) and they propose a personalized ranking algorithm for results re-ranking. Similarly, [9] combines results categorization and personalized IR techniques to provide a personalized

concept-based search mechanism over LOD. In this approach, results are categorized based on a conceptual ontology, UMBEL (umbel.org) [8]. Then, based on the user's click through data on concepts and results, the search results are adapted; for instance they apply concept re-ranking, results re-ranking, query expansion or concept suggestions [9]. The common aspect of [6, 7, 9] is that all approaches receive initial search results from a search engine and apply clustering (concepts) as the basis of the personalization. Although these approaches [6, 7, 9] individually evaluated the impact of their concept-based personalization, they only focused on retrieval precision and ignore the user aspects of such personalization. In this paper, we focus on this problem and evaluate personalized concept-based search against the ranked lists in terms of task assistance and user satisfaction. Then, we provide some preliminary conclusions.

2 Linked Open Data Search Approaches

Current LOD search engines use ranked lists to present search results [1] [2]. In this approach, results order depends on relevancy to the user query. Ranked lists work well if the user has a specific information need. However, by the nature of the LOD, resources are structured and often resources contain scattered data that is linked to other resources. This makes it even harder, for unfamiliar users, to search and explore the LOD. Faceted search [4] and personalized search [9] approaches are proposed to overcome this problem. In this work, we use [9] to compare with the ranked lists. In particular, results are grouped based on their concept categorizations [8]. Then, initial results are presented with no adaptation, where result categories are ranked according to relevance to the query. The following personalization is applied when the user interacts with the search system: (i) When a user selects a concept for exploration; all concepts are re-organized according to their semantic and syntactic similarity. In addition, within the selected concept, more relevant results are included using results re-ranking and query expansion as well as relevant concepts are suggested for results exploration. (ii) When the user clicks on a result, within the interacted concept, immediately personalization is applied. Such as, relevant results and concepts are added by query expansion and results re-ranking according to last *N* clicks of the users [9].

3 Evaluations and Analysis

3.1 Experimental Setup

Dataset: Our evaluation is based on information seeking tasks in a tourism domain, particularly "tourism in Killarney Ireland". We selected a tourism domain since it suits well for data gathering and informational queries as well as users can issue specific queries as they learn the topic. We use the benchmark dataset as explained in [9].

Ranked List vs Personalized Concept-based Search: As a comparison, a purpose-built non-adaptive ranked list search system was created. For a fair comparison, the ranked list (i) uses the same underlying indexing and retrieval models, (ii) operates across the same dataset, and (iii) results are ranked according to relevancy to the query as with the personalized concept-based search. The only difference is that the personalized system categorizes and adapts the results to the click through data [9].

Tasks: The search tasks were inspired from Google popular search queries in our domain. In order to test different types of queries, 4 search tasks were prepared with varying level of specificity, such as from fact finding tasks to information gathering tasks [10]. The questions were deliberately very specific, this enabled users to decide if they were satisfied to complete the task. Task 1 and 3 contained a mixture of open-ended and fact finding questions. Task 2 and 4 contained open-ended questions.

Experiment: We used 32 participants from School of Computer Science and Statistics of Trinity College Dublin (1 master student, 20 PhD students, 9 post-docs and 2 academics). For a fair comparison, users were divided into two groups, such as Group A and Group B. Group A users performed either Task 1 or Task 3 using the ranked list first and then performed Task 2 or Task 4 using the personalized search. Similarly, Group B users performed Task 1 or Task 3 using the personalized system first and then performed Task 2 or Task 4 using the ranked list. Thus, all tasks were equally tested and both systems equally (and anonymously) presented as the first system.

3.2 Results and Analysis

Task Assistance: It is desirable that a system requires users to invest the least amount of effort in order to find relevant information as quickly as possible. The results from the task completion times revealed that the personalized concept-based search outperformed the ranked list with an average of 6.50 (m:ss) versus 10.48. Moreover, t-tests confirm that the results are indeed significant for each task (for Task 1 $p=0.037$, for Task 2 $p=0.047$, for Task 3 $p=0.03$ and for Task 4 $p=0.003$) as shown in Figure 1. It is also shown that Task 2 and especially Task 4 took considerably longer to complete using the ranked list compared to the personalized search. The reason for this could be that both Task 2 and Task 4 contained open-ended questions, compared to Task 1 and Task 3. Similarly, users formulated fewer queries across all tasks using the personalized system (Figure 1). Again t-tests confirm that the results are significant (for Task 1 $p=0.045$, for Task 2 $p=0.031$, for Task 3 $p=0.017$ and for Task 4 $p<0.001$). On average, users issued 6.46 queries using the ranked list to complete the tasks compared to average of 3.03 queries for the personalized system. Another aspect of task assistance is the number of viewed pages. It is desirable that the search system provides the best resources in the top results, thus users require few page views. The results showed that users consistently required more page views across all tasks using the ranked list (average of 10.56 versus 5.34 page views as shown in Figure 1). T-tests also confirms the significance (Task 1 $p=0.048$, Task 2 $p=0.048$, Task 3 $p=0.038$ and Task 4 $p=0.013$). This gain was mainly obtained by the personalization; results re-ranked as well as more relevant results were automatically pushed on top of the search list using category/results re-ranking and query expansion techniques using our system.

User Satisfaction: The findings were backed up by post-questionnaires as shown in Figure 2. These results are indeed statistically significant with $p<0.001$ for Q1-Q12. In addition, users strongly recognized and valued categorization (Q8 - with an average of 4.43 versus 1.78) and personalization aspects (Q12 - with an average of 3.96 versus 1.68). Moreover, personalized concept-based search achieved an average SUS usability score of 88.90 compared to 75.15 of the ranked list. This is an interesting finding, especially considering the familiarity of users with the traditional ranked lists. These

results are even more encouraging with personalization features like re-ranked lenses, re-ranked results and category suggestions, users thought that the personalized concept-based search was easy to use and better than the ranked list in terms of usability.

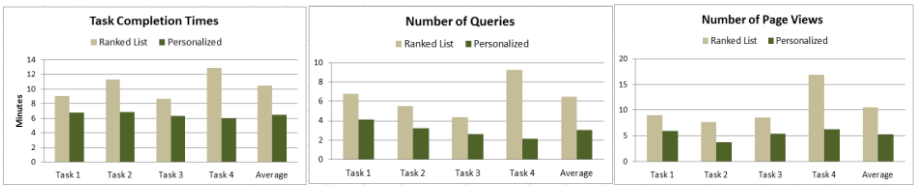
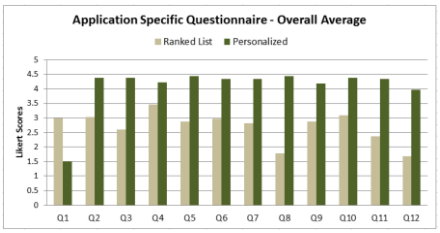


Figure 1. Task completion times, number of queries and page views across tasks



Q1: I had to search a lot before I found interesting content.
Q2: I spent less time querying and more time browsing.
Q3: I was less exposed to irrelevant content.
Q4: I did well on tasks.
Q5: The results structure and content was helpful to solve the tasks.
Q6: I am satisfied with the system performance, guidance and assistance.
Q7: I found the presentation of the search results helpful.
Q8: I found the categorization and grouping of search results helpful.
Q9: The result structure and content matched my expectations accurately.
Q10: The content composition generated by the system was easy to navigate.
Q11: I felt guided to relevant results.
Q12: The system guided me towards more personally relevant content.

Figure 2. Users' perceptions of search, result presentation and personalization

Analysis: Personalized concept-based search enabled users to be efficient and effective at performing both information gathering and fact finding tasks as shown by task completion times, number of issued queries and viewed pages. Questionnaires backed up these findings. Results are indeed statistically significant for all metrics.

Acknowledgements. This research is supported by the Science Foundation Ireland (Grant 12/CE/I2267) as part of CNGL (www.cngl.ie) at Trinity College Dublin.

4 References

1. Delbru, R., S., Campinas, G., Tummarello: Searching Web Data: an Entity Retrieval and High-Performance Indexing Model. Journal of Web Semantics, 10, 33-58, 2012.
2. D'Aquin, M., E., Motta, M., Sabou, S. Angeletou, L., Gridinoc, V., Lopez and D., Guidi: Toward a New Generation of Semantic Web Applications. IEEE Intelligent Systems, 2008.
3. Tummarello, G., R. Cyganiak, M. Catasta, S. Danielczyk, R. Delbru and S. Decker: Sig.ma: live views on the Web of Data, Journal of Web Semantics, 8(4), 355-364, 2010.
4. Heim, P., T. Ertl and J. Ziegler: Facet Graphs: Complex Semantic Querying Made Easy, Extended Semantic Web Conference, LNCS, 6088, 288-302, 2010.
5. Hogan, A., J. Umbrich, A. Harth, R. Cyganiak, A. Polleres, S. Decker: "An empirical survey of Linked Data conformance. Journal of Web Semantics, 14, 14–44, 2012.
6. Ferragina, P., and A., Gulli: 2005. A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering. In: International conference on World Wide Web, 801-810, 2005.
7. Lee, H.C., and Borodin, A: Cluster Based Personalized Search. Algorithms and Models for the Web-Graph, 5427, 167-183, 2009.
8. Sah, M., and V. Wade: A Novel Concept-based Search for the Web of Data using UMBEL and a Fuzzy Retrieval Model. In: Extended Semantic Web Conference, 7295, 103-118, 2012.
9. Sah, M., and V. Wade: Personalized Concept-based Search and Exploration on the Web of Data using Results Categorization. In: Extended Semantic Web Conference, 7882, 532-547, 2013.
10. User study tasks can be accessed from https://www.scss.tcd.ie/melike.sah/tasks.html
11. Jansen, B.J., D.L., Booth, A. Spink: Determining the User Intent of Web Search Engine Queries. International Conference on World Wide Web, 1149-1150, 2007.