PivotE: Entity-Oriented Exploratory Search on Knowledge Graphs

Jun Chen
chenjun2013@ruc.edu.cn
Renmin University of China

Xueran Han
hanxueran@ruc.edu.cn
DEKE Lab (MOE), Renmin University of China

Yueguo Chen
chenyueguo@ruc.edu.cn
DEKE Lab (MOE), Renmin University of China
National Engineering Laboratory of Big Data System Software(Beijing Institute of Technology)

Jiaheng Lu
jiaheng.lu@helsinki.fi
University of Helsinki

Giulio Jacucci
giulio.jacucci@helsinki.fi
University of Helsinki

Tuukka Ruotsalo
tuukka.ruotsalo@helsinki.fi
University of Helsinki

Abstract

Entity-oriented exploratory search in knowledge graphs can enrich information access by presenting similar entities and their relevant exploration pointers to fulfill users' information needs, which potentially supports a particular set of entity search and recommendation tasks. However, less attention has been devoted to the user interfaces for supporting users to explore the knowledge graphs effectively and efficiently. In this paper, we focus on the user study of a prototype system called PivotE for entity-oriented exploratory search in knowledge graphs. It applies a state-of-the-art method for recommending entities and their relevant information as exploration pointers, which assists users to learn about the properties of entities and guides them to explore the knowledge graphs in different aspects. It also allows users to manipulate entities and exploration pointers directly to express their information needs beyond the keyword-based search, which facilitates users to reformulate queries. Extensive task-based user experiments indicate that, in terms of task results, search behaviors and user experiences, PivotE achieves improved performance significantly.

Index Terms: Exploratory Search—Knowledge Graph—Explainable Entity Recommendation—User Interface Prototype System;

1 Introduction

Entity-oriented exploratory search can enrich information access by presenting entities instead of documents instantly, as well as offering relevant exploration pointers to fulfill users' information needs [1]. The goal of exploratory search aiming at better answering users' information needs beyond lookup activities, which involves multiple interactions that take place over perhaps very long periods of time and may return results that are critically assessed before being integrated into the users' knowledge such as learning and investigation activities [11]. However, exploratory search is not trivial to be implemented, because users are often unfamiliar with the information space, as well as unclear to express their information needs, which result in that the query formulation evolves iteratively as they become more familiar with the search context and their information needs [3]. Therefore, during a long-term search process, particular supports in formulating queries, learning about unfamiliar information space and identifying possible search directions are very important to users. Currently, traditional search systems based on knowledge graphs focus on supporting users to iteratively formulate queries for better addressing semantic search [4, 7, 8]. However, less attention has been devoted to the user interfaces for supporting users to explore the knowledge graphs in different aspects, which has three main challenges: 1) conventional user interactions are constrained by the keyword-based input, which confronts users with challenges in expressing their information needs clearly and reformulating queries rapidly in the unfamiliar information space; 2) millions of entities are connected by thousands of relations in knowledge graphs, which confronts systems with challenges in recommending relevant entities and semantic features effectively and efficiently; 3) conventional search systems often force users to narrow down the information space continually, which confronts users with the challenges of switching across the multi data domains freely in the information space.

In this paper, we introduce an entity-oriented exploratory search prototype system called PivotE [5], which have been accepted by VLDB 2019. This paper mainly discusses the interactive interfaces of the exploratory process. The key idea is that, given the typed-query keywords as an initial query, it returns entities with respect to these keywords and their relevant semantic features, which explains the semantic correlation among the entities in one view. It not only assists users to learn about the properties of entities in many aspects, but also guides them to further explore the information space. Besides, it supports horizontally and vertically explore different data domains by allowing users to manipulate entities and semantic features directly. In such a way, the gap between unstructured keyword-search and structured knowledge graphs can be bridged via the entities and semantic features. The main contributions of this paper can be summarized as follows:

- we propose the design principles as guidelines for implementing the prototype system that can support entity-oriented exploratory search in knowledge graphs effectively and efficiently.
- we conduct extensive task-based user experiments, in terms of task results, search behaviors and user experiences, our design principles can facilitate the entity-oriented exploratory search process in knowledge graphs.

2 Design Study

We first study design challenges, and then propose design principles to implement our prototype system.

2.1 Design Challenges

We identify the main challenges of entity-oriented exploratory search in knowledge graphs as follows:

1. How to express information needs semantically and formulate queries rapidly.

Activities considered as related to exploratory search are very diverse and hard to define in a consolidated way. But unlike the basic lookup activities, they usually take place in areas that are unfamiliar to users and are characterized by the frequent need to reformulate queries. However, conventional user interfaces for information retrieval in knowledge graphs are constrained by the typed-query keywords, which confronts users with challenges in expressing their

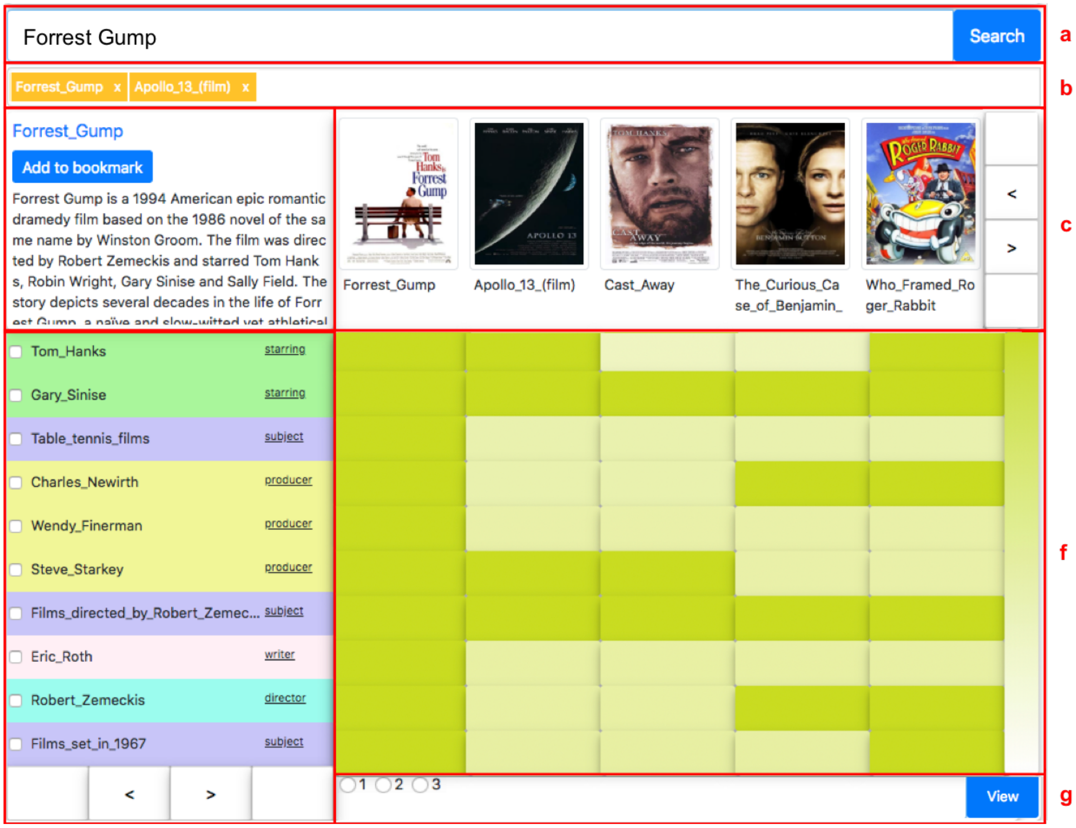


Figure 1: User interface of PivotE, the main workspace is divided in four areas: the query area (see Fig. 1-a, b and g), the entity recommendation area (see Fig. 1-c), the entity presentation area (see Fig. 1-d), the semantic feature recommendation area (see Fig. 1-e), the explanation area (see Fig. 1-f) [5]

information needs semantically, and then slow down the iterative search process of query reformulation.

2. How to learn about the search context and identify the possible search directions.

In current search systems, users are often forced to collect and learn the cues from the intermediate results by themselves, instead of focusing on collecting and learning from relevant information. To address such problems, the recommendation of relevant exploration pointers is a means to assist users to learn about the search context and identify the possible search directions. However, millions of entities are connected by thousands of relations in knowledge graphs, which may result in that the recommended information are too diverse to recommend. The lack of explanations for such recommendation may confuse users, and then lower acceptance and trust towards the results returned by the system.

3. How to go through a long-term search process and explore the information space in different aspects.

Exploratory search can transcend multiple query iterations during the long-term search process, which may result in that users are often lost in the unfamiliar information space. Besides, traditional search systems on knowledge graphs focus on supporting users to iteratively formulate queries for semantic search, which often forces users to narrow down information space continually, and then lower their enthusiasm to further explore the other data domains. Therefore, it is important that the system could support the searches over time, as well as support users to explore the information space in different aspects.

2.2 Design Principles

Based on the above mentioned design challenges, we propose design principles as guidelines to implement our prototype system as follows:

A. Utilizing entities and semantic features to express information needs and formulate queries for addressing challenge 1.

To overcome the limitation of traditional user interfaces, we support users to express their information needs by using entities and semantic features (see Fig. 1-c and e). For instance, users can express the query intent *"Find films starring Tom Hanks"* by specifying the semantic feature *Tom_Hanks:starring*, as well as express the query intent *"Find films similar to Forrest Gump"* by simply specifying the entity *Forrest_Gump*. In such a way, users can not only narrow the information space in different aspects, but also deeply investigate similar entities in the same data domain.

To support query formulation rapidly, entities and semantic features could be used to formulate queries (see Fig. 1-b), either individually or combined, to gain a new set of entities and their relevant semantic features as search results. Besides, the possibility to input keywords is still necessary for some situations, for instance, when the system fails to make the proper suggestions or when specifying an initial query (see Fig. 1-a). Therefore, the system should support an alternative way to instantiate a new search session.

B. Recommending relevant semantic features of entities and supplying the semantic correlation between them as explanations for addressing challenge 2.

To learn about the information space and foster the understanding of entities, we should help users get rid of discovering cues from the unstructured information (see Fig. 1-d). To address this issue, relevant semantic features can be recommended as the properties of entities and exploration pointers for further exploration (see Fig. 1-e). However, if the mechanics and reasoning of the recommendation algorithms can be communicated to users in the right way, it can improve acceptance and trust towards the system (see Fig. 1-f). For instance, if the system explains the semantic correlation

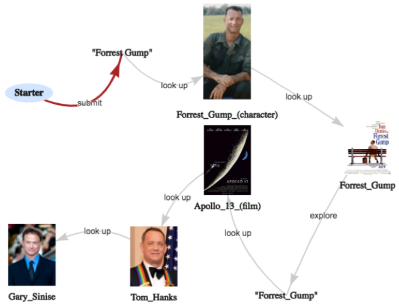


Figure 2: Visualization of search behaviors

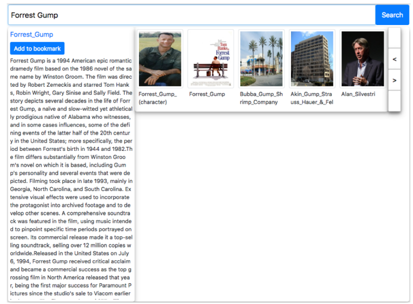


Figure 3: User interface of Baseline 1

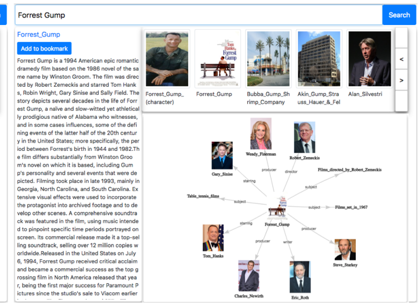


Figure 4: User interface of Baseline 2

between *Forrest_Gump* and *Apollo_13_(film)* are that both of them are performed by *Tom_Hanks* and *Gary_Sinise*, users may better learn about the search context, and then identify a more reasonable search direction for further exploration (e.g., further exploring the films performed by *Tom_Hanks* by specifying the semantic feature *Tom_Hanks:starring*).

C. Organizing queries in a timeline for traceback, as well as supporting free switches across different data domains for addressing challenge 3.

To support the activities during a long-term search process, we organize queries and visualize search behaviors in a timeline to remind the position of users, as well as support users to revisit the previous results immediately (see Fig. 1-g and Fig. 2). In such a way, the persistence of search context improves exploratory search by fostering trials without fear of losing the current work, and supporting information comparison by revisiting historical queries. Moreover, to help users get rid of a specified data domain and facilitate them to explore the information space in different aspects, we allow them to freely switch across different data domains relevant to the current one, rather than confusedly leap to irrelevant ones (see Fig. 1-e). For instance, users can vertically explore relevant entities (e.g., *Tom_Hanks*) from the neighboring data domains (e.g., *Actor*) of the current one (i.e., *Film*), via the relevant semantic features (e.g., *Tom_Hanks:starring*).

3 User Study

We designed an experiment to verify whether the above design helps the user to explore knowledge graphs. In our experiments, we apply the DBpedia 2014 as the dataset, and design 3 tasks to evaluate the performance of PivotE and two baselines using the measures by considering the task results, search behaviors and user experiences.

3.1 Dataset

The English version of DBpedia 2014[1] is applied as the dataset in our experiments. It describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including 1.45 million persons, 0.74 million places, 0.41 million creative works, 0.24 million organizations, 0.25 million species, 6,000 diseases, etc.

3.2 Baselines

We implement two baselines. Baseline 1 supports keyword-based input and returns relevant entities matching to the given keywords (see Fig. 3). Compared to Baseline 1, Baseline 2 also visualizes the semantic features of an entity, and supports further exploration by clicking the entities in the semantic features such as *Tom_Hanks* (see Fig. 4). These two baselines apply the same dataset and ranking algorithms for recommendation to compare them with PivotE fairly. All these factors aimed to create the baselines that allow us to focus the evaluation solely on our design principles.

[1]https://wiki.dbpedia.org/data-set-2014

3.3 Tasks

In order to encourage the participants to interact with the systems, we design 3 tasks that require the participants to select entities meeting complex information needs, i.e., learning or investigation, rather than trivial matches to the keywords, by considering the diversity of what they selected within 5 minutes as follows:

1. select 5 entities relevant to the topic *"JFK"*.
2. select 5 films similar to *"Forrest Gump"*.
3. select 5 films performed by the actors or actresses that have ever cooperated with the director *"Yimou Zhang"*, but not directed by him.

3.4 Participants and Procedure

We recruited 18 students from our universities. For the testing procedure, we follow a within-subjects experiment design, counter-balanced by changing the order of the 3 tested systems and tasks (i.e., Latin Square Design). In such a way, each task will be done by 6 participants in different orders. Before starting the tasks, users will receive a detailed instruction about how to use the systems. Each task will give users a short guideline of what type of entities they need to select. Following these guidelines, they should select entities as their answers to the given tasks, and provide their confidence of what they selected. After using a system, they will be given two short questionnaires, then move to the next task.

3.5 Measures

The evaluation measures of the user experiments are designed considering the following 3 factors: task results, search behaviors, and user experiences.

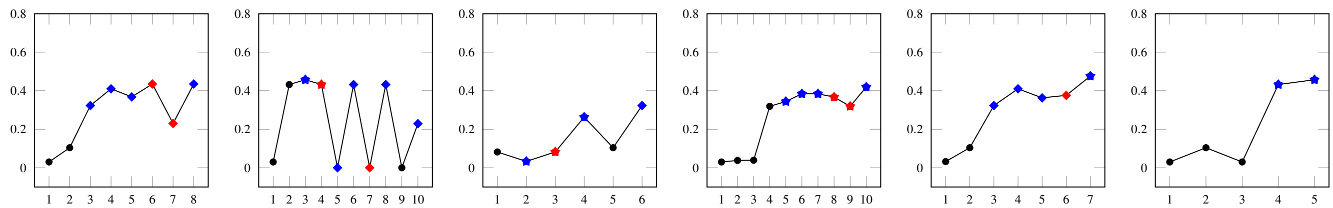
Task Result Evaluation In order to compare the performance of different systems for finishing the tasks, we need to evaluate the quality of the results using different systems. Before evaluation, we first create the ground truths for each task by pooling the retrieved entities of every query-response during this task. Experts are asked to assess the relevance of the retrieved entities on a binary scale (i.e., 0=irrelevant, 1 relevant). After that, the task-level results are measured by the means and standard deviations of the relevance and confidence of the entities selected by different participants.

Search Behavior Evaluation In order to investigate users' search behaviors, we have recorded search trail parameters for each system during each task using a method resembling White's [10] as follows:

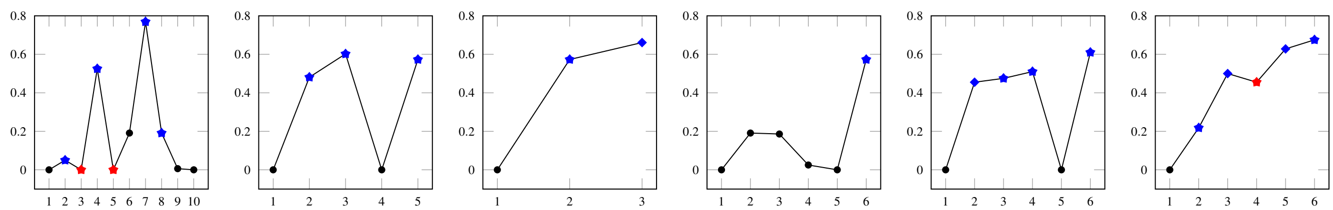
1. #queries: the number of queries;
2. #queries (keywords): the number of keyword-based queries;
3. #queries (entities): the number of queries that manipulate entities to reformulate the query;
4. #queries (semantic features): the number of queries that manipulate semantic feature to reformulate the query;
5. #queries (selections): the number of queries that select entities as the answers to the task;
6. #queries (revisits): the number of queries that have been submitted earlier during the task;

Table 1: The results of search trail analysis of 3 systems. Results are reported as the means (M) and standard deviations (SD) of the search trails of different participants during each task.

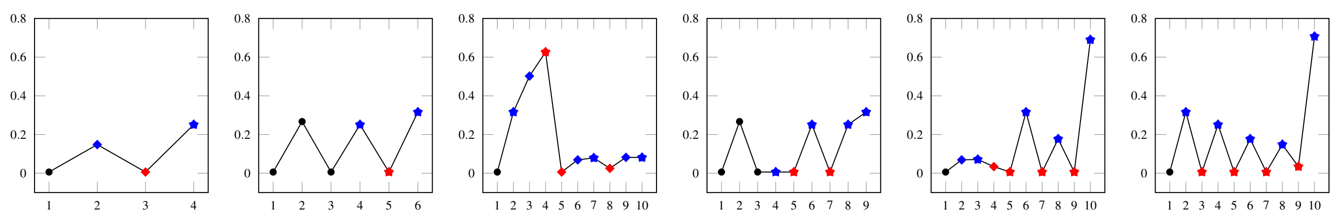
Search Trail Parameters	Task 1						Task 2						Task 3					
	Baseline 1		Baseline 2		PivotE		Baseline 1		Baseline 2		PivotE		Baseline 1		Baseline 2		PivotE	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
#queries	2.33	1.75	5.83	3.31	**8.00**	3.29	5.17	3.31	4.17	2.14	**6.50**	4.32	4.00	2.28	6.83	2.14	**8.50**	3.08
#queries (keywords)	2.33	1.75	2.00	1.10	**2.83**	0.98	**5.17**	3.31	1.34	0.52	2.33	1.75	**4.00**	2.28	3.17	2.56	1.83	0.98
#queries (entities)	0.00	0.00	**3.83**	2.17	3.33	3.67	0.00	0.00	**2.84**	0.98	1.67	0.82	0.00	0.00	**3.66**	1.03	1.67	1.86
#queries (semantic features)	0.00	0.00	0.00	0.00	**1.83**	1.04	0.00	0.00	0.00	0.00	**2.50**	1.78	0.00	0.00	0.00	0.00	**5.00**	2.03
#queries (selections)	1.50	0.84	3.50	1.52	**3.70**	0.89	2.50	1.38	2.50	1.05	**3.67**	1.51	2.17	1.33	3.67	1.63	**4.00**	1.55
#queries (revisits)	0.17	0.41	0.33	0.52	**1.17**	1.94	0.50	0.55	0.33	0.82	**1.33**	2.34	0.00	0.00	1.17	1.17	**2.33**	1.51
#queries/min	2.17	1.92	2.45	0.69	**2.96**	1.11	3.75	1.88	2.50	2.86	**3.04**	1.98	2.75	2.11	3.54	2.98	**4.55**	2.48
time cost (min)	1.90	1.72	2.28	1.21	**2.78**	0.69	1.54	0.96	3.14	2.91	**3.24**	2.26	2.16	1.88	**3.78**	3.85	2.98	1.37



(a) The MAP results of the sequential query-response of task 1



(b) The MAP results of the sequential query-response of task 2



(c) The MAP results of the sequential query-response of task 3

Figure 5: The query-level effectiveness results of PivotE by different participants. Results are reported as the MAP of the sequential query-response during each task. The vertical axis represents the map value of the query, and the horizontal axis represents the query sequence, where the black filled circles refer to keyword-based queries, the blue and red filled diamonds (or stars) refers to the queries that manipulate the entities (or semantic features) to formulate the queries, the blue (or red) ones refer to add (or delete) an item to (or from) the current query.

7. #queries/min: the number of queries;

8. time cost (min): the time cost of finishing the task.

Besides, we also evaluate the quality of the results derived from every query-response during each task, by the metrics considering the precision and recall, including P@k (Precision@k), MRR (Mean Reciprocal Rank) and MAP (Mean Average Precision) [6].

User Experience Evaluation In order to find which system is most beneficial for users to finish the tasks, we need to analyze the subjective feedbacks of the participants after using different systems via two questionnaires:

1) the standard SUS (System Usability Scale) questionnaire, consisting of 10 questions. It is widely used and validated for measuring perceptions of system usability [2];

2) the standard ResQue questionnaire, consisting of 15 questions. It is a widely used and validated for measuring perceptions towards the recommender systems [9].

4 RESULTS

In this section, we analyze the results of user experiments based on the above proposed measures. In details, we are going to answer the main research questions as follows:

- **RQ1**: Do our design principles assist users to better finish the tasks?
- **RQ2**: Do our design principles facilitate users' search behaviors to explore the information space?
- **RQ3**: Do our design principles improve users' satisfactions towards the system usability and the recommended results?

4.1 Task Result Evaluation

We first compare the performance of different systems for finishing the tasks. Generally, according to the mean relevance for each task as illustrated in Tab. 2, our prototype system PivotE and the Baseline 2 substantially improve the performance than Baseline 1, which shows that the keyword-based search cannot effectively support exploratory search in knowledge graphs. Compared to Baseline 2, PivotE is still favorable because it supports users to express their information needs

Table 2: The task-level relevance of 3 systems. Results are reported as the means (M) and standard deviations (SD) of the relevances of the entities selected by different participants during each task.

Tasks	Baseline 1		Baseline 2		PivotE	
	M	SD	M	SD	M	SD
1	0.57	0.15	0.63 (+12%)	0.20	**0.83** (+47%)	0.20
2	0.33	0.24	0.50 (+50%)	0.28	**0.67** (+100%)	0.10
3	0.47	0.27	0.73 (+57%)	0.10	**0.80** (+71%)	0.18
All	0.46	0.12	0.62 (+37%)	0.12	**0.77** (+68%)	0.09

more flexible by utilizing entities and semantic features, as well as the explanations among them to make users better understand the search context during the tasks. In such a way, users can freely explore the knowledge graphs in different aspects, which fosters the methods for understanding towards the unfamiliar information space, and then assist users to make more accurate decisions.

To observe how our design principles affect users' decisions during each task, we compute the task-level confidences of 3 systems as illustrated in Tab. 3. Generally, according to the mean confidence for each task, we find the results of PivotE are much higher than Baseline 1 and Baseline 2, which shows that the presentation of structured and semantic information along with the entities, i.e., semantic features and explanations, improves users' confidences for finishing the tasks. When looking into the distribution of confidence (i.e., SD) for each task, we find that the results of our prototype system PivotE and Baseline 2 are clustered more tightly around the mean than Baseline 1. A possible explanation for such results is that when users have the semantic features (or combined with the explanations) along with the entities, they are more confident and agreeable to make the decisions.

Table 3: The task-level confidence of 3 systems. Results are reported as the means (M) and standard deviations (SD) of the confidences of the entities selected by different participants during each task.

Tasks	Baseline 1		Baseline 2		PivotE	
	M	SD	M	SD	M	SD
1	3.77	0.73	4.17 (+11%)	0.46	**4.70** (+25%)	0.11
2	3.17	0.64	4.10 (+29%)	0.30	**4.73** (+46%)	0.17
3	3.33	0.72	4.20 (+26%)	0.49	**4.73** (+42%)	0.33
All	3.42	0.31	4.16 (+20%)	0.05	**4.71** (+38%)	0.02

Overall, these above results illustrate that our design principles can improve the effectiveness for finishing the tasks, as well as improve users' confidences and agreements to make the decisions during the tasks, which answers the research question **RQ1** from the perspective of task results.

4.2 Search Behavior Evaluation

In order to observe how the design principles affect users' search behaviors during the tasks, we first analyze the search trails of 3 systems. As illustrated in Tab. 1, according to the means of search trails for different systems during each task, we find that the users in PivotE submit more queries (i.e., #queries) and spend more time (i.e., time cost (min)) than the ones in Baseline 1 and Baseline 2 in most cases. Further more, we find there is a high ratio of queries associated with entities and semantic features (i.e., #queries (entities) and #queries (semantic features)) for Baseline 1 and PivotE, which shows that users are motivated to utilize the entities or semantic features to express their information needs. Moreover, the direct manipulations of entities and semantic features also foster query reformulation (i.e., #queries/min).

When looking into the search trail parameter #queries (selections), we find that the users in Baseline 2 and PivotE are more willing to submit more queries to select entities as their answers to improve the diversity of what they selected. This is because users in Baseline 1 need to learn the cues from the intermediate results by themselves, which slows down the iterative search process and their willingness to further explore the information space. Compared to Baseline 2, PivotE is still favorable, which illustrates that our design principles can inspire users to attempt more possible search directions and therefore facilitate them to further explore the information space in different aspects.

For the remaining parameter of search trail, i.e., #queries (revisits), we find that users sometimes need to revisit some historical queries during the tasks, especially for PivotE that supports users to easily revisit a historical query. This is reasonable because when exploring the unfamiliar information space, users sometimes need to trace back to a specified position, and then attempt a new search branch. The introduction of query management and search behavior visualization improves exploratory search by fostering trials without fear of losing during the long-term search process. It also support information comparison by revisiting historical queries immediately.

Table 4: The query-level effectiveness results of 3 systems. Results are reported averagely for the P@10, P@20, MRR and MAP of each query-response by different participants during the task including revisits.

Systems	Task 1			
	P@10	P@20	MRR	MAP
Baseline 1	0.20	0.18	0.35	0.07
Baseline 2	0.32 (+59%)	0.23 (+34%)	0.60 (+71%)	0.14 (+95%)
PivotE	**0.47** (+13%)	**0.37** (+11%)	**0.73** (+11%)	**0.25** (+24%)
Systems	Task 2			
	P@10	P@20	MRR	MAP
Baseline 1	0.17	0.14	0.21	0.06
Baseline 2	0.24 (+48%)	0.26 (+90%)	**0.34** (+62%)	0.11 (+83%)
PivotE	**0.27** (+65%)	**0.29** (+11%)	0.28 (+30%)	**0.14** (+13%)
Systems	Task 3			
	P@10	P@20	MRR	MAP
Baseline 1	0.20	0.23	0.17	0.12
Baseline 2	0.30 (+50%)	0.24 (+7%)	**0.43** (+15%)	0.16 (+33%)
PivotE	**0.43** (+11%)	**0.35** (+53%)	0.32 (+87%)	**0.24** (+103%)
Systems	All Tasks			
	P@10	P@20	MRR	MAP
Baseline 1	0.19	0.18	0.25	0.08
Baseline 2	0.29 (+52%)	0.25 (+36%)	**0.46** (+87%)	0.14 (+63%)
PivotE	**0.39** (+11%)	**0.34** (+87%)	0.44 (+80%)	**0.21** (+15%)

In order to observe effectiveness results of each query-response during each task in more details, we first generate query-level effectiveness results of PivotE by different participants as illustrated in Fig. 5, the results are reported as the MAP of the sequential query-response during each task. According to the MAP derived from different participants during each task, the results of the queries reformulated by entities or semantic features often achieve significant improvements than the results of the keyword-typed queries. We also compute the query-level effectiveness results including revisits in Tab. 4. Since pivotE needs to guide the user to rebuild the query and browse data domain multiple times to implement complex queries, the RR[2] value of the first few queries in the query sequence will be lower, so the MRR value of pivotE in Tab. 4 is smaller than baseline2. But the performance of PivotE is significantly better than the others, because it allows users to select entities by considering the diversity of their data domains, which shows that our design principles facilitate users' search behaviors, and users can take advantages from such search behaviors.

Overall, these results illustrate that our design principles lead to more active search behaviors, with more queries per minute and a high ratio of queries reformulated by entities and semantic features. Besides, we find the semantic features and explanations along with entities are very beneficial for users to understand the search context, and to attempt more possible search directions via rapid query

[2]The reciprocal of the first correct result position in each query is recorded as RR, and MRR means the average of RRs.

reformulation, which answers the research questions **RQ2** from the perspective of user behaviors.

4.3 User Experience Evaluation

In order to analyze the subjective feedbacks of the participants after using 3 systems, we generate the results of SUS and ResQue questionnaires derived from different participants for different systems. As illustrated in Fig. 6, the results of the SUS questionnaire are M=71.22; SD=7.78 for PivotE, Mean=56.89; SD=6.37 for Baseline 1 and M=68.58; SD=9.88 for Baseline 2. A paired t-test shows that PivotE significantly outperforms the two baselines, revealing higher usability for PivotE. Besides, the results of the ResQue questionnaires are also favorable for our prototype system, i.e. M=60.67; SD=6.98 for PivotE, M=28.78; SD=5.67 for Baseline 1 and M=43.94; SD=6.92 for Baseline 2. A paired t-test shows that PivotE significantly improves than two baselines, revealing that the users using PivotE have higher satisfaction towards the recommended results. These results subjectively answer the research question **RQ3** from the perspective of user experiences.

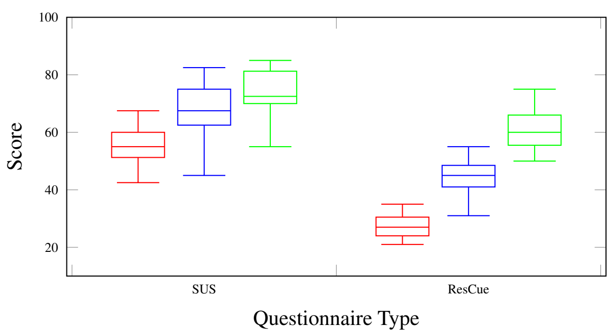


Figure 6: The results of SUS and RecCue questionnaires of 3 systems. Results are reported as the scores of different participants, where the red box plot (i.e., the left one) refers to the scores of Baseline 1, the blue box plot (i.e., the middle one) refers to the ones of Baseline 2, and the green box plot (i.e., the right one) refers to the ones of our prototype system PivotE. A paired t-test shows that PivotE significantly outperforms the two baselines using $p < 0.05$.

5 Discussions and Conclusions

Exploratory search aims at better answering users' information needs beyond lookup activities, which involve multiple interactions that take place over perhaps very long periods of time and may return results that are critically assessed before being integrated into the user's knowledge [11]. In order to support exploratory search, particular supports in formulating queries, learning about unfamiliar information space and identifying possible search directions are very important to users.

5.1 Methodological Contributions

We propose novel design principles for implementing the prototype system that can support entity-orient exploratory search in knowledge graphs effectively and efficiently. We design task-based user experiments to evaluate the performance of the prototype system using the measures by considering the task results, search behaviors and user experiences.

5.2 Answers to the Research Questions

Extensive task-based user experiments are studied to evaluate the performance of our prototype system PivotE. According to the task-level results, the query-level results, the search trail analysis and the subjective feedbacks, we find the benefits and answer the research questions as follows: 1) our design principles can improve the effectiveness to finish the tasks, and they improve users' confidences and agreements to make the decisions during the tasks,; 2) our design principles lead to more active search behaviors, with more queries per minute and a high ratio of queries reformulated by the entities and semantic features, and they improve users' understanding towards the search context, and then attempt more possible search directions via rapid query reformulation; 3) our design principles bring a better system usability and user satisfaction for assisting users to deal with the complex tasks.

5.3 Summary and Future Work

We design and implement a novel prototype system called PivotE for entity-oriented exploratory search in knowledge graphs. It applies a state-of-the-art method for recommending entities and their relevant information as exploration pointers, which assists users to learn about the properties of entities and guides them to explore the knowledge graphs in different aspects. It also allows users to manipulate entities and exploration pointers directly to express their information needs beyond the keyword-based search, which facilitates users to reformulate queries. Extensive task-based user experiments demonstrate that, in terms of task results, search behaviors and user experiences, our design principles are found to facilitate entity-oriented exploratory search in knowledge graphs. In the future, for evaluation measures, we are extending our relevance assessments towards novelty and diversity, as well as including comparisons with more tasks, user interfaces, entity search, and entity recommendation algorithms. This will ensure evaluation motivated by even more realistic information needs on a continuously developing benchmark.

6 Acknowledgments

Yueguo Chen is the corresponding author. This work is supported by National Key Research and Development Program (No. 2018YFB1004401) and the National Science Foundation of China under grants U1711261, No. 61472426.

References

[1] K. Balog, E. Meij, and M. de Rijke. Entity search: Building bridges between two worlds. In *SEMSEARCH, Raleigh, North Carolina, USA*, pp. 9:1–9:5. ACM, 2010.

[2] J. Brooke. Sus: A quick and dirty usability scale. 189, 11 1995.

[3] R. Capra, G. Marchionini, J. S. Oh, F. Stutzman, and Y. Zhang. Effects of structure and interaction style on distinct search tasks. In *JCDL*, pp. 442–451, 2007.

[4] K. Elbedweihy, S. N. Wrigley, P. D. Clough, and F. Ciravegna. An overview of semantic search evaluation initiatives. *J. Web Sem.*, 30:82–105, 2015.

[5] X. Han, J. Chen, J. Lu, Y. Chen, and X. Du. Pivote: Revealing and visualizing the underlying entity structures for exploration. *Accepted by PVLDB'19*.

[6] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008.

[7] E. Meij, K. Balog, and D. Odijk. Entity linking and retrieval for semantic search. In *WSDM, New York, New York, USA*, pp. 683–684. ACM, 2014.

[8] R. Neumayer, K. Balog, and K. Nørvåg. When simple is (more than) good enough: Effective semantic search with (almost) no semantics. In *ECIR*, pp. 540–543, 2012.

[9] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *RecSys*, pp. 157–164, 2011.

[10] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *WWW*, pp. 21–30, 2007.

[11] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.