EntEXPO: An Interactive Search System for Entity-Bearing Queries

Xitong Liu, Peilin Yang, and Hui Fang

University of Delaware, Newark, DE, USA
{xtliu,franklyn,hfang}@udel.edu

Abstract. The paper presents EntEXPO, a search system that aims to improve the search experience for entity-bearing queries. In particular, the system exploits the entities and their relations in the context of query to identify a list of related entities and leverage them in an entity-centric query expansion method to generate more effective search results. Moreover, EntEXPO displays the related entities along with the search results to allow search users explore entity relationship to further refine the query through an interactive interface.

Keywords: entity centric, interactive search.

1 Introduction

Entities have been playing an important role in the information seeking process. Many Web search queries are related to entities. For example, Pound et al. [1] found that more than 50% of queries in a sample of Web queries are related to entities. In the meantime, significant efforts have been put to build knowledge bases such as Freebase, DBpedia and YAGO, which contain valuable information about entities and their relations. Therefore, it would be interesting to study how to leverage these knowledge bases to further improve the search experience of Web search users for entity-bearing queries.

In this paper, we describe our efforts on building a novel search system, i.e., EntEXPO, which exploits entity relationships in the context of queries. The system first identifies entities from a query and then retrieves a list of related entities based on the information from both the document collection and knowledge base. The related entities are used for (1) entity-centric query expansion [2,3], which expands original queries with the related entities to improve the search quality; and (2) entity-centric query reformulation, which allows search users to interact with the system by manually adjusting the weight for each related entity.

Our demonstration plan includes the following three parts. First, we will demonstrate the effectiveness of search by using entity-centric query expansion method. In particular, we will conduct side-by-side comparison between these results with those generated by using existing pseudo relevance feedback methods and highlight the differences. Second, we will explain that the generated entity-relation graph can help users understand the relations among entities in

M. de Rijke et al. (Eds.): ECIR 2014, LNCS 8416, pp. 784–788, 2014.

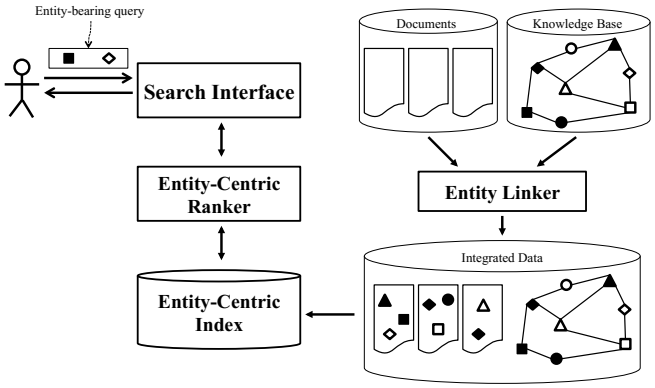

Fig. 1. System Architecture of EntEXPO

a better way and enable them to explore the information space more effectively than existing query suggestion results. Finally, we will discuss how the related entities and their relations can be used to reformulate the query by changing the weights of the related entities in the weight panel.

2 System Description

Figure 1 shows the architecture of EntEXPO, and we now describe the major components of the system in more details.

Entity Linker: Entities and their relationships exist in both the document collection and knowledge base. The information from the knowledge base is often structured, while those from the data collection is not. The main functionality of *entity linker* is to extract the entity information from the document collection using Named Entity Recognizer and link them with the corresponding entries in the knowledge base. This would allow us to leverage the entity related information in a more systematic way.

Entity-Centric Index: In addition to the traditional term-based inverted index, we also need build an entity-based inverted index on the integrated data to accelerate the process of finding documents that mention some entities. Entity-centric index is particularly useful when finding related entities for a given query.

Entity-Centric Ranker: The entity-centric ranker performs two tasks. It first retrieves a list of related entities for the query based on the entity relations in the integrated data. With the related queries, it could either automatically expand the original query using the entity-centric approach proposed in our previous studies [2,3] or allow users to reformulate the query by manually adjusting the weight for each related entity.

Search Interface: Besides the search box and document retrieval list as in the regular Web search result page, the search interface has two more components:

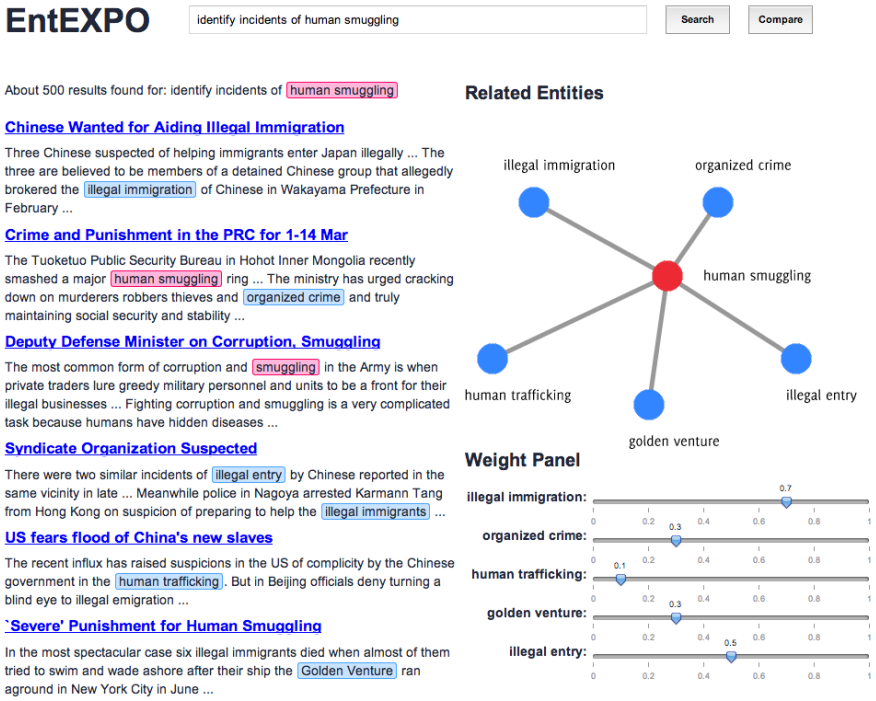

Fig. 2. Search interface of EntEXPO

related entity graph and entity weight panel. Figure 2 shows the search interface of EntEXPO. The query is "identify incidents of human smuggling". We notice that "human smuggling" is identified as a named entity and denoted in pink color in the top of left column. The right column shows the related entity graph, which visualizes the relation between the query entity and related entities[1] including "illegal immigration", "organized crime", "human trafficking", "golden venture" and "illegal entry". All the related entities are denoted in blue color in the snippets of search results. If the user clicks on one related entity in the graph, the documents which contain the related entity will be highlighted accordingly to help her understand why these documents are retrieved. If the user clicks on the bar connecting the query entity and a related entity, documents mentioning both entities will be shown in the result list to help her understand the relations between them. The entity weight panel is located beneath the related entity graph, in which each related entities is associated with a weight slider. By default each slider shows the weight suggested by EntEXPO automatically. The user can adjust the weight of each related entity individually to reformulate the query based on her inspection of document retrieval list.

[1] At most the top 5 related entities will be shown. If no related entity is found, the related entity graph and entity weight panel will remain blank, and search results will be retrieved based on default model without query expansion.

3 Demonstration Plan

In the demo, we use the documents from TREC 2004 Robust Track as the document collection and the English version of DBpedia 3.7 as the knowledge base. The document collection contains 528,155 documents, while the DBpedia contains 3.77 million entries with 400 million facts. Search users are allowed to type their own queries. Alternatively, they can also select a query from the official query set from TREC 2004 Robust track. The major advantage of selecting TREC queries is that users can immediately see the effectiveness of retrieval methods since the judgments for these queries are available.

We plan to demonstrate three main features of EntEXPO as follows:

- **Better retrieval effectiveness:** Users can try out their own queries and go through the search results by themselves, as shown in Figure 2. More importantly, we will also provide a comparison interface to allow users to compare the results of EntEXPO with those using existing pseudo relevance feedback methods side by side. If user selects a TREC query, the interface will display the effectiveness of both systems and highlight the difference between their search results.
- **Entity-centric information exploration:** EntEXPO displays an entity-relation graph along with the document retrieval results. The graph contains the entities from the query as well as those related entities with respect to the query. The information about these entities and relations (by clicking the nodes or edges) allows users to learn more information about the query. This feature is expected to be particularly useful for exploratory search, where users need more guidance to navigate the information spaces in order to formulate appropriate queries.
- **Interactive query formulation.** By leveraging the entity weight panel, users can reformulate the query by rewarding or penalizing the related entities. Upon receiving the new results from the entity-centric ranker, user may continue adjusting the weight to explore the parameter space iteratively, or stop when satisfying results are found. The benefits are two-fold: (1) it saves user's efforts from manually reformulating query by changing the query in the search box and conducting search once again. (2) it provides more flexibility to reformulate the query with precise control of term weighting than adding or removing terms alone.

4 Conclusion

We propose and develop a novel search system, i.e., EntEXPO, for entity-bearing queries. It exploits entity relationship to improve search accuracy as well as providing an interactive interface to explore the information space surrounded by the related entities and reformulate the query. We believe that EntEXPO could, in particular, benefit users whose information seeking tasks are exploratory and entity-related.

References

1. Pound, J., Mika, P., Zaragoza, H.: Ad-hoc Object Retrieval in the Web of Data. In: WWW, pp. 771–780 (2010)
2. Liu, X., Fang, H., Chen, F., Wang, M.: Entity Centric Query Expansion for Enterprise Search. In: CIKM, pp. 1955–1959 (2012)
3. Liu, X., Chen, F., Fang, H., Wang, M.: Exploiting Entity Relationship for Query Expansion in Enterprise Search. Information Retrieval (2013) (in press)