

Query refinement: type aware approach

Zahra Taherikhonakdar
University of Windsor
taherik@uwindsor.ca

Aaditya Pradipbhai Parekh
University of Windsor
parekh23@uwindsor.ca

ABSTRACT

Returning relevant documents based on the user's initial query is a major challenge for search engines. The user's initial query is ambiguous for search engines as it is usually short, has a different meaning, and the query term does not represent the user's search intent. One of the existing solutions to overcome this limitation is query refinement in order to enhance the retrieval performance by improving document ranking and obtaining additional relevant documents. The user's query has different aspects/types. Knowing the different types of the query would result in reducing initial query ambiguity and better understating the user's search intent. In this proposal, we consider different query types and evaluate the different query refinement techniques based on them to select the appropriate refining methods. This would result in information retrieval(IR) methods improvement and the user's search satisfaction.

KEYWORDS

Query Refinement, Query Types, Information retrieval

1 INTRODUCTION

Considering query types would help search engines to better understand the user's search intent. Border [1] divided web search queries into three categories, namely: navigational, informational, and transactional queries. If the user wants to find specific information on a specific topic, the query is named informational query. These queries usually result in a set of documents rather than just one suitable result. With navigational queries, the user wants to find a certain webpage that he/she is already familiar with or thinks that such a webpage exists. Searching for a person or company homepage is considered a navigational query. Transactional queries result in Web sites where doing further interaction is necessary. Downloading music or file, purchasing a product are examples of transactional queries.

Santos et al. [4] considering informational and navigational query types using the TREC 2009 data set [2] to infer the intent of the query. They focus on selecting suitable retrieval models for different query types to improve the diversification in results. Researchers try to define different types of queries in different domains to increase IR performance and also increase diversity in retrieving results. Vallet et al. [3] proposed the method to predict important types based on informational queries. They introduce the top three query types in the informational query category, namely: country, organization, events. They claim that this would increase the performance of the informational query search.

2 MOTIVATION

In this research proposal, we propose to consider query types in query refinement to improve IR performance. We use the TREC 2009 data set [2] and use different query refinement methods (e.g. tagmee, wordnet) for all three query types in the data set. Using metrics (e.g. NDCG) we would evaluate the IR performance by our approach. Based on the evaluation, the appropriate query refinement method is considered for each query type. To be specific, our proposal tries to find the best possible answer for this question "how to expand the user's initial query given the type of this query". We expected selecting an appropriate query refinement method for each query type improves the IR performance in terms of the relatedness of retrieved documents.

2.1 Motivating Example

Assume that the user's initial query is 'led zeppelin' (an English rock band). The user could have different intentions like downloading songs, the homepage of bands, and searching biography of the band. These intentions represent transactional, navigational, and informational query types, respectively. Knowing the query intention, we could select the appropriate query refinement method to retrieve related documents and enhance the user's search satisfaction. For instance, we could conclude that the refinement of the 'led zeppelin' would be improved by applying a query refinement method suitable for the transactional query.

3 PROBLEM DEFINITION

Given a set of queries $Q = q_1, q_2, \dots, q_n$, we classify them into the different types of query $\mathcal{T} = t_1, t_2, t_3$. Given the different query suggestion methods $\mathcal{M} = m_1, m_2, \dots, m_n$, we select the top three methods for each query types based on the evaluation metric like NDCG.

4 TEAM JUSTIFICATION

Zahra and Aaditya are both interested in the area of web search. Although we will participate in all parts of the project, the research and writing section would be by Zahra since her research domain is in this area, and implementing code would be by Aaditya since he has Python programming knowledge.

REFERENCES

- [1] Andrei Broder. 2002. A taxonomy of web search. In *ACM Sigir forum*, Vol. 36. ACM New York, NY, USA, 3–10.
- [2] Charles L. Clarke, Nick Craswell, and Ian Soboroff. 2009. *Overview of the trec 2009 web track*. Technical Report. WATERLOO UNIV (ONTARIO).
- [3] David Vallet and Hugo Zaragoza. 2008. Inferring the most important types of a query: a semantic approach. In *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval*. 857–858.
- [4] Eric Zhan, Jagjeet Singh, AI Argo, Yisong Yue, and Andrew Hartnett. [n.d.]. The Argoverse Trajectory Retrieval Benchmark. ([n. d.]).