Query refinement: type aware approach*

Zahra Taherikhonakdar
University of Windsor
taherik@uwindsor.ca

Aaditya Pradipbhai Parekh
University of Windsor
parekh23@uwindsor.ca

ABSTRACT

The goal of this research is to improve query refinement considering query types. In general, in the proposed method, the TREC2009 Million Query dataset [3] is used as the input of the ReQue [2] to investigate the different query expanders' performance given different query types. In this report, we present the results of the proposed method evaluation.

KEYWORDS

Query Refinement, Query Types, Information retrieval

1 EXPERIMENTAL SETUP AND EVALUATION

1.1 Dataset

In our experiments, we use a publicly available TREC 2009 Million Query Track [3]. It consists of approximately 40,000 queries collected from a large search log. About 400 queries out of 40,000 queries were categorized into the following query types: navigational, informational-close, informational-open, resource, and advice. The aim of navigational queries is to find a specific URL. Queries which categorized into informational-close and informational-open categories. Their aim is to find the direct (unambiguous) information and undirect information (seek all available information about the topic), respectively. Queries in the resource category are also called transnational queries with the aim of, for example, downloading music from a website. Advice queries find the ideas regarding a general question [1].

1.2 Setup

The main part of the methods is done by ReQue [2]. The ReQue is a tool to examine different expander methods on the input queries and generate refined queries which improve the information retrieval task. The whole experiment consists of two steps:

1.2.1 Preprocess input data. In this step, we preprocess the TREC2009 Million Query dataset and its relevance judgment to be accepted as an input for the ReQue.

1.2.2 Running the ReQue. In this step, we run the ReQue on the preprocessed dataset. The ReQue requires three inputs: 1- the datasets of queries and their relevance judgments. 2- an information retrieval method which in this experiment, we select the BM25 method. 3- an evaluation metric which MAP is chosen for this experiment. The results of the ReQue would be refinements of initial queries which enhance the performance of the information retrieval task and the expanders which produce these refined queries. We use Tagmee, Glove, and LovinsStemmer as query expanders methods in ReQue.

---

*https://github.com/ZahraTaherikhonakdar/Eval-QR-type-aware-approach

1.3 Metrics

As the performance of the information retrieval task with trec2009 data set is run and evaluated by the ReQue, we set the ReQue to use BM25 [5] as an information retrieval task and use MAP (mean average precision) [4] as our evaluation metric.

1.4 Evaluation

In this section, we evaluate the results based on the specified query types. We build a matrix of query expanders and query types and add one if the expander improves the query in specific query types. For instance, if the Tagmee expander method could improve the input query in navigational query type, we add one value to the specific cell based on each improvement.

Table 1 contains the example of refined queries by given query expanders.

Query Expander	Initial Query	Refined Query
Tagmee	cuttle fish bone	cuttle fish bone Cuttlefish
Glove	premedication for dental procedures	premedication dentists dentistry required
Lovins	pinecrest tropical gardens	pinecrest trop garden

Table 1: Refined queries by different query expanders

Figure 1 shows the matrix. Columns C1, C2, and C3 represent Tagmee, Glove, and Lovins expanders, respectively. Rows 1, 2, 3, 4, and 5 represent informational-close, informational-open, navigational, resource, and advice, respectively.
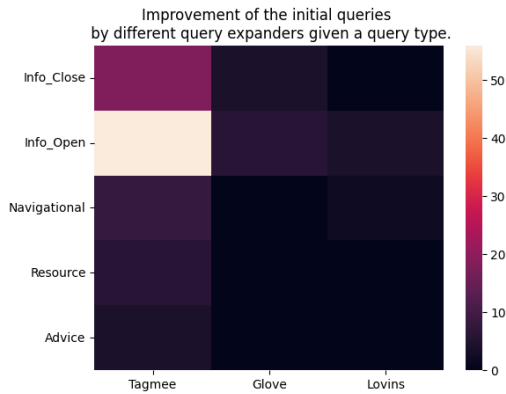

	C1	C2	C3
1	18	4	0
2	56	6	4
3	8	0	2
4	6	0	0
5	4	0	0

Figure 1: The Matrix.

In this experiment, we have 78 informational-close, 213 informational-open, 59 navigational, 42 resource, and 14 advice queries.

For better representation of the results, we draw the heat-map of the given matrix. Figure 2 shows the heat map.

Figure 2: The Heat Map.

As figure 1 and figure 2 show, Tagmee improves all query types but improves informational-open queries the most. The Glove only improves informational-open and informational close, respectively. The Lovins improves informational-open and navigational queries. All query types improve most by the Tagmee expander.

Zero in each cell means that the expander could not improve any query in the specific query type or that query type could not be improved by the specific query expander. For instance, quires in the advice type could not be improved by the Glove expander, or the Lovins could not improve the queries in the informational-close type.

One reason that fewer or zero queries in navigational, resource, and advice categories could be improved by given query expanders could the fewer quantity of them are exist in the input.

For future work, we are going to have more quires of each type and examine more query expanders to come up with comprehensive and reliable results.

REFERENCES

[1] Ben Carterette, Virgiliu Pavlu, Hui Fang, and Evangelos Kanoulas. 2009. Million Query Track 2009 Overview.. In *TREC*. Citeseer.

[2] Mahtab Tamannaee, Hossein Fani, Fattane Zarrinkalam, Jamil Samouh, Samad Paydar, and Ebrahim Bagheri. 2020. Reque: a configurable workflow and dataset collection for query refinement. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 3165–3172.

[3] TREC. 2010. *TREC 2099 Million Query Track*. https://trec.nist.gov/data/million.query09.html

[4] Wikipedia contributors. 2022. Evaluation measures (information retrieval) — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Evaluation_measures_(information_retrieval)&oldid=1077036698 [Online; accessed 3-May-2022].

[5] Wikipedia contributors. 2022. Okapi BM25 — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Okapi_BM25&oldid=1082128332 [Online; accessed 3-May-2022].