# BJUT at TREC 2017: Tasks Track

**Lupu Li**[1], **Guangyuan Zhang**[1], **Zhen Yang**[1]✉
1. College of Computer Science, Faculty of Information, Beijing University of Technology, China
✉yangzhen@bjut.edu.cn

### Abstract

We generally evaluate the quality of task based information retrieval system from two aspects: task understanding and task completion. To evaluate the quality of task understanding, we need a list of sub-tasks that provide a complete coverage of tasks for each query. Finding a way to get a list like that for every query is our target. So, this paper explored the viable approach to achieve it and analyzed experimental results.

## Introduction

Unlike traditional search engines only focused on serving the best results for a single query, the task based search engines try to understand the reasons that might have motivated the user to submit that query. The Tasks track is aimed at devising techniques to evaluate task based information retrieval systems ability to understand the tasks (Verma et al. 2016). The Tasks track is divided into two parts: Task understanding and Task completion.

This paper focused on Task understanding, it ask for a ranked list of keyphrases that represent the set of all subtasks for a query, while avoiding redundancy. The quality of the ranked list of keyphrases is evaluated using diversity-aware metrics. We mainly makes use of two keyphrase sources:

- Google suggestion API and Bing suggestion API
- Generated by combing the initial query and keywords extracted from documents.

The method details will be covered in Section 2.

## Ranking System Framework

The set of candidate keyphrases is ranked by scoring. The score of a keyphrase q was generated by the initial query is:

$$S(q) = S_s(q) + S_k(q) \tag{1}$$

The score of keywords from web search engine suggestion marked as $S_s$, and the score of keywords from keyword extraction marked as $S_k$. We used the query suggestions from the Google and Bing suggestion API, and give them all 10.0 points as the highest score, because the web search engine suggestion is the most available source.

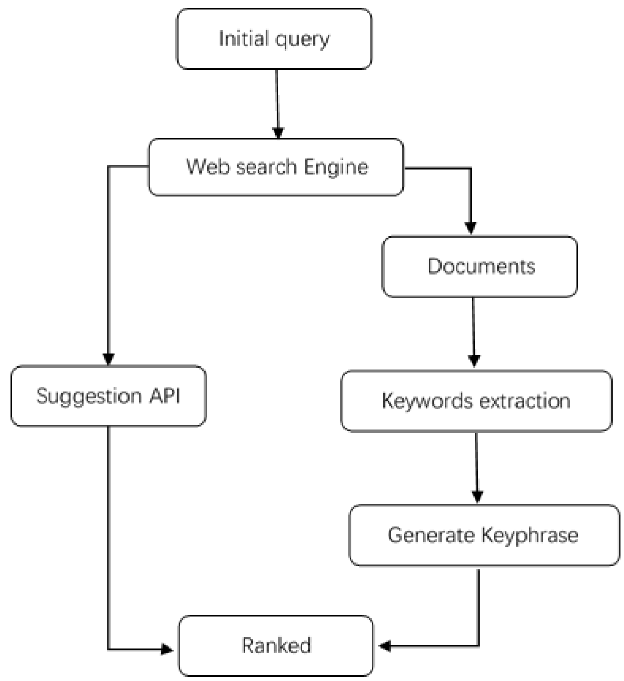



Figure 1: Ranking System Framework.

### Keyword extraction

The other major source of our candidate keywords is extracting keywords from documents. The source of documents in $BJUT_1$ (RUN ID) is the top 10 results from google search engines, and it in $BJUT_2$ is the top 10 results from $Clueweb12$ document collection.

Before extracting keywords, we removed the stop words and numbers, because they are meaningless in Generating keyphrases. We extracted keywords from each document using the ED keyword extraction algorithm(Yang et al. 2013), this algorithm use difference between the intrinsic mode and extrinsic mode to extract the highly relevant keywords. and the score S2 is offered by this algorithm. The score of keyword means the relevance between keyword and document. The higher the score, the higher the relevance. We removed the keywords with a score less than 0, and if the documents were from web search engines, we would remove the HTML elements.

In the final, we get a set of keywords, we used two different ways of combing a keyword $k$ and the initial query q: (i)adding $k$ as a suffix to $q$; (ii) adding $k$ as a suffix to the last entity in $q$. Such a removal strategy is motivated by the fact that in English most of the usual (non-adjectival) renements are syntactically performed as an addition to the right.(Garigliotti and Balog 2016)

### Scoring keyphrase

We ranked keyphrases by their scores. The score of a keyword k extracted from document d is generated by:

$$S_k(k) = \sum_{i=1}^{n} s(k|d_i)s(d_i|q) \tag{2}$$

- $s(k|d_i)$: expresses the score of keyword k extracted from document. it is gived by keyword extraction algorithm, The score of keyword means the relevance between keyword and document. The higher the score, the higher the relevance.
- $s(d_i|q)$: expresses the score of document $d_i$ generated by the initial query q. it was gived by Clueweb12 document collection. if the document was from web search engine, we set it to 1 points.

Combining Eq(1) and Eq(2), we will have the final score of keyphrases. So, the graphical representation of the model is depicted in Figure 1.

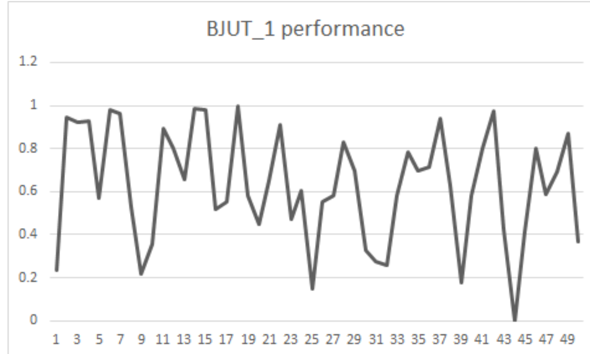
## Experimental Results

We submitted the following two runs: $BJUT_1$ and $BJUT_2$, The results of our runs is shown in Figure 2 and Table 1. We don't have more statistics about other participators because of only have two teams submitted run this year. So we can't compare the results with other teams who participated this year. About the details of two runs is as follow:

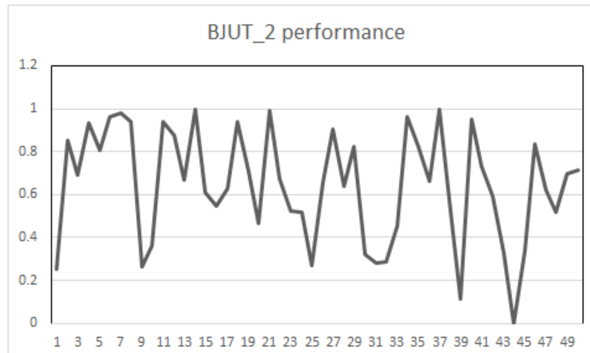
- $BJUT_1$: we used the web search engine suggestions with the highest priority in the ranking, and we chosen top-10 documents from results of web search engine as the source of keyword extraction. we only used keywords with $S_k(q)$ more than 0.
- $BJUT_2$: This approach is essentially the same as $BJUT_1$, but the source of documents is Clueweb12 document collection.

### Results Analysis

We compare the two runs to each other. We can see that the two results are generally similar, Because both of them used the web search engine suggestion API, and it contributed the majority of the score. Due to the $BJUT_1$ used documents from Web search engine ,and it have some interference from HTML elements, the scores of $BJUT_1$ is lower than the scores of $BJUT_2$. But the differences of two scores is very small.




(a) 1




(b) 2

Figure 2: alpha-nDCG@20 Performances.

Table 1: Experimental mean Performances.

| Run | nERR-IA@20 | alpha-nDCG@20 |
|---|---|---|
| $BJUT_1$ | 0.536024 | 0.627844 |
| $BJUT_2$ | 0.561675 | 0.643954 |

## Conclusions

We have described our participation in the TREC 2017 Tasks track, and it will be the last Task track because of the few participators. We explored the viable approach to generate phrases which represent sub-tasks of initial query, and we experimented the two sources of documents what be used to extract keywords.

In future work, we will find other sources of keywords, methods for combining keywords, and methods for generating phrases.

## Acknowledgments

This research was supported by the Data Mining & Security Lab and professor Yang from Beijing University of Technology.

## References

Garigliotti, D., and Balog, K. 2016. The university of stavanger at the trec 2016 tasks track. In *TREC*.

Verma, M.; Yilmaz, E.; Mehrotra, R.; Kanoulas, E.; Carterette, B.; Craswell, N.; and Bailey, P. 2016. Overview of the trec tasks track 2016. In *TREC*.

Yang, Z.; Lei, J.; Fan, K.; and Lai, Y. 2013. Keyword extraction by entropy difference between the intrinsic and extrinsic mode. *Physica A: Statistical Mechanics and its Applications* 392(19):4523–4531.