IMU Experiment in IR4QA at NTCIR-8

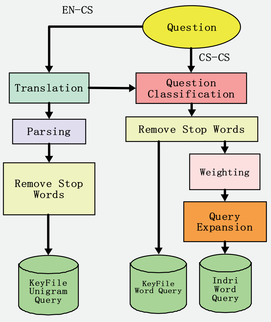
Xiangdong Su, Xueliang Yan, Guanglai Gao, Hongxi Wei
School of Computer Science
Inner Mongolia University
Hohhot, China 010021
Email：csggl@imu.edu.cn

Introduction

This poster describes the key points of the experiments we conducted for the NTCIR-8 IR4QA subtask. The basic assumption we based on is that a document is more relevant to a query if it is supported by multi-evidence returned from different methods. So we create three indexes of different formats and then retrieve from them to examine the documents returned. If a document appeared more than one time, it will be ranked higher than the others, although, as we will see later, the rule is not just so simple since we will assign weights to different indexes. To avoid query drift, we use external resource to perform query expansion. We conduct our experiments by using the Lemur Toolkit (http://www.lemurproject.org) and the official evaluation results show that our method can do help improving the performance.

QUERY PROCESSING

We do query processing first since we need to retrieve from different index format.

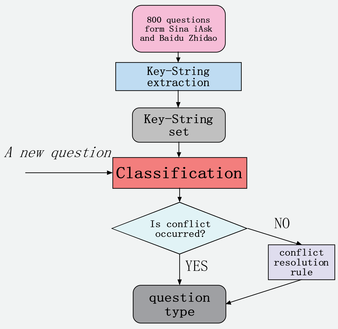


TRANSLATION

Before the process of retrieval, we need to convert the questions to a suitable format. In EN-CS Run, we use the Google Translation Services to translate English questions into Chinese ones. Other treatments to English topics are same as the Chinese ones.

QUESTION CLASSIFICATION

For different question types, we will do specific query expansion later. So we must classify the questions into different types now.
We classify the questions into 9 types based on the occurrence of the key-strings together with some conflict rules. The key-strings are extracted from 800 questions from Sina iAsk (http://iask.com/) and Baidu Zhidao (http://zhidao.baidu.com/) by statistical approach.



Key-String examples for DATE type questions

Key-String: 哪年，哪月，哪天，何时，何年,何月，何日，什么时候，什么时间，什么时辰，多会儿，具体时间，生日，诞辰，纪念日

An example of conflict resolution rule for BIOGRAPHY and PERSON type questions

R1: If except "谁" ("who"), "谁是"("Who is"), "是谁" ("Who is") and punctuation, there is only NR tag in the question after been tagged, the question is BIOGRAPHY type, and otherwise it belongs to PERSON type.

WEIGHTING

The importance of terms in a query is not the same. Term weight reflects the discriminative power of the term in query. A proper weighting scheme could enhance the retrieval effectiveness. Since Lemur does not support weighting operation, there's no weighting in the process of KeyFile-Unigram-Query and KeyFile-Word-Query generation. In experiments, we only use term weighting in Indri-Word-Query.
Term-weighting schemes assign weights to terms relying on how useful they are likely to be in determining the relevance of a document . We use the following scheme for term weighting:
(1) For questions of BIOGRAPHY and DEFINITION, there is no term weighting.
(2) For questions of RELATIONSHIP, we do weighting according to the collection frequency of two objects in the question. When the frequencies of the two objects are not in the same scale, we increase the weight of low frequency object. Otherwise, there is no weighting.
(3) For questions of other types, we increase the weights of the most important two key words, which are judged by Hailiang API (www.hylanda.com) .

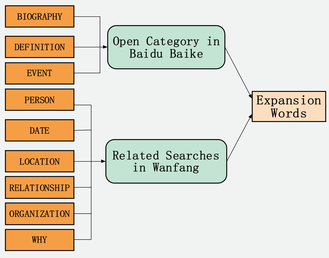
Performance comparison on weighting in the training experiments with the NTCIR-7 IR4QA CS test collection

Weighting	Indri Word Query		
	mean AP	Q-measure	nDCG
No	0.5435	0.5396	0.7438
Yes	0.5542	0.5484	0.7494

Legend: Yes represents result using term weighting, No represents result without term weighting

QUERY EXPANSION

We perform the query expansion according to the question types mentioned earlier.
For questions belong to BIOGRAPHY , DEFINITION or EVENT, we use Open Category in Baidu Baike (http://baike.baidu.com) to do the expansion. Otherwise, we adopt the Related Searches provide by Wanfang (http://www.wanfangdata.com.cn).



Open Category in Baidu Baike

entry	Open Category
何大一	台湾，人物，博士，科学院院士
李宇春	明星，华人明星，歌手，超级女声，影响力人物
老龄化社会	社会保障，社会保险，社会问题， 老人，人口研究
319枪击案	台湾

Related Search in Wanfang

KeyTerm	藏历	重合
ty_max	历法,时令节气,时轮历,藏历	不动点,匹配,相交,重合
ty_mid		主位,主语,平移,旋转角度,移动变换,超凸度量空间,非紧性测度
ty_min		g-凸空间,双务合同,定性对策,履行,截口,抽象经济

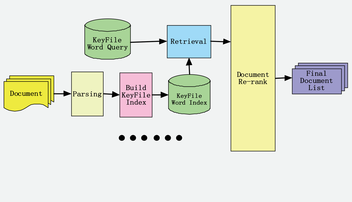
Performance comparison between results with and without Open Category expansion in the training experiments with the NTCIR-7 IR4QA CS test collection

Expansion	Indri Word Query		
	MAP	Q-measure	nDCG
No	0.5837	0.5579	0.6820
Yes	0.6060	0.5800	0.7180

Legend: No represents result without expansion, Yes represents result with expansion

INDEXING, RETRIEVAL AND RE-RANK

Our basic idea is that documents supported by more evidences are more likely to relevant to the query, and we considered the results retrieved from different indexes as the evidences.
Lemur Toolkit was adopted to create three indexes of different types: KeyFile-Unigram-Index, KeyFile-Word-Index and Indri-Word-Index.
After retrievaling, we re-rank the result documents by a simple interpolating algorithm.



Re-rank algorithm

Description: This algorithm is designed to implement the document re-rank task according to the returned document lists from three indexes mentioned above.
(1)score normalize:
For each returned document list from above three indexes
  For each Di in the returned document list
    Score normalize(Di);
(2)compute score:
  For each Di appeared in one of the returned document list
    Score interpolating(Di);
(3)sort documents:
  Sort(score list);

The score function used to Re-rank is as follows:

$$Score(D_i) = \alpha Score_{KU}(D_i) + \beta Score_{KW}(D_i) + \gamma Score_I(D_i)$$

Where:
α，β，γ： The interpolating coefficients of the scores of D_i in the corresponding returned document lists.
$Score_{KU}(D_i)$: The score of D_i in the returned list from KeyFile-Unigram-Index. If D_i is not in the returned list from KeyFile-Unigram-Index, it equals to zero.
$Score_{KW}(D_i)$: : The score of D_i in the returned list from KeyFile-Word-Index. If D_i is not in the returned list from KeyFile-Word-Index, it equals to zero.
$Score_I(D_i)$: : The score of D_i in the returned list from Indri-Word-Index. If D_i is not in the returned list from Indri-Word-Index, it equals to zero.

Performance of single index and their combination in the CS-CS training experiments with the NTCIR-7 IR4QA CS test collection

Result	CS-CS						
Measure	①	②	③	①+②	①+③	②+③	①+②+③
AP	0.4354	0.5424	0.5435	0.5685	0.5600	0.5483	0.5735
Q-measure	0.4445	0.5389	0.5396	0.5693	0.5564	0.5441	0.5738
nDCG	0.6736	0.7378	0.7438	0.7710	0.7579	0.7458	0.7733

① KeyFile-Unigram-Index
② KeyFile-Word-Index
③ Indri-Word-Index
+ means the combination of them

Official Evaluation Results

Official Evaluation Results together with the parameters adopted

Measure	IMU-CS-CS-01-T	IMU-CS-CS-02-T	IMU-CS-CS-03-T	IMU-EN-CS-01-T
α	0.04	0.16	0.25	0.04
β	0.16	0.24	0.25	0.16
γ	0.80	0.60	0.50	0.80
mean AP	0.4266	0.4114	0.4032	0.3184
mean Q	0.4628	0.4480	0.4394	0.3540
mean nDCG	0.6761	0.6580	0.6575	0.5720

CONCLUSION

A simple weighting scheme has been used in our IR system, which brings a slight improvement to performance. We exploit Open Category in Baidu and Related Searches in Wanfang to solve the vocabulary mismatch. Our experiments show that combination of indexes using different formats can improve retrieval performance. They also confirm that combination of indexes using different index units helps the retrieval performance. So we believe index combination could obtain a more satisfactory result in general.

Contact Information

School Of Computer Science
Inner Mongolia University
Hohhot, 010021
Inner Mongolia Autonomous Region, China
Email: csggl@imu.edu.cn
Tel:+86-471-4992341