Homework 3

0 Collaboration and Originality

1. Did you receive help of any kind from anyone in developing your software for this assignment (Yes or No)? It is not necessary to describe discussions with the instructor or TAs.

   NO

2. Did you give help of any kind to anyone in developing their software for this assignment (Yes or No)?

   NO

3. Did you examine anyone else's software for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

   NO

4. Are you the author of every line of source code submitted for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

   YES

5. Are you the author of every word of your report (Yes or No)?

   YES

Instructions

1 Experiment 1: Baselines

1.1 Experimental Results

	Indri			
	BOW		**Query Expansion**	
	Reference System (*given*)	**Your System (Exp-1a)**	**Reference System (Exp-1b)**	**Your System (Exp-1c)**
P@10	0.6800	0.7000	0.7400	0.7100
P@20	0.6450	0.6300	0.6750	0.6500
P@30	0.5600	0.5300	0.5667	0.5467
MAP	0.1934	0.1806	0.2066	0.1877
win:loss	N/A	Loss	Win	Loss

1.2 Parameters

retrievalAlgorithm=Indri

Indri:mu=1500

Indri:lambda=0.01

fb=Indri

fbDocs=10

fbTerms=10

fbMu=0

fbOrigWeight=0.5

1.3 Discussion

Overall Discussion:

Query Expansion has large improvement comparing to search only on original query. Sometimes, users do not know what exactly they want. Sometimes, they do not know how to express it precisely. Sometimes, they do not want to waste time to carefully design a query. And it turns out, we can learn from the data to infer what the information need is.

Example Examination:

- Example1

711 has the largest improvement after applying query expansion. MAP improves from 0.19 to 0.30. Original query 711:Train station security measures. The provided expansion: 711: #wand(0.0031 measure 0.0026 amtrak 0.0024 security 0.0019 train 0.0019 transit 0.0017 station 0.0011 police 0.0011 rail 0.0010 transportation 0.0010 all)

As we can see, some same words are retrieved. These word increases the weight for those "common" and "important" word in the original query. Some words with similar concepts are retrieved. For example, "rail", "transportation". These words could effectively match more relative documents, in case synonyms of original words are used instead of original words. Finally, we can see some words like "police". This is very interesting. As we can see "police" has to do with the "security" in the original query. Although the user type in "security" in the query, it does not have to do with the "police". However, based on the data from relative documents, our model extends the original meaning of the query. This could be helpful when users expect the search engine should have some ability to inferring the meaning of the query instead of only matching or understanding the query.

- Example2

Not all query expansion has improvement. Some of the query gets worse result. 730 decreases 0.02 points in MAP. The original query: 730: Gastric bypass complications. The query expansion: 730: #wand(0.0030 obese 0.0028 gastric 0.0023 weight 0.0023 surgery 0.0020 bypass 0.0017 patient 0.0014 complication 0.0011 health 0.0011 loss 0.0010 severe)

This query is very technical. Therefore, words like gastric are added. In this case, the preciseness of the query is very important. The terms inferred by the model could add errors to the query.

Robustness: When query improves the search quality, it usually improve it a lot. When it decorates the performance, it is relatively small.

2 Experiment 2: The number of feedback documents

2.1 Experimental Results

	Indri BOW, Reference System (*given*)	**Query Expansion, Reference System Initial Results**				
		Feedback Documents				
		10 (Exp-2a)	**20 (Exp-2b)**	**30 (Exp-2c)**	**40 (Exp-2d)**	**50 (Exp-2e)**
P@10	0.6800	0.7100	0.7600	0.7100	0.7200	0.7200
P@20	0.6450	0.6500	0.6200	0.6500	0.6150	0.6300
P@30	0.5600	0.5467	0.5467	0.5467	0.5400	0.5300
MAP	0.1934	0.1877	0.1866	0.1877	0.1869	0.1874
win:loss	N/A	Loss	Loss	Loss	Loss	Loss

2.2 Parameters

retrievalAlgorithm=Indri

Indri:mu=1500

Indri:lambda=0.01

fb=false

fbTerms=10

fbMu=0

fbOrigWeight=0.5

2.3 Discussion

Overall discussion:

There is a sweet point for the parameter fbDocs. If the number of documents is too small. We could not have enough candidates for the expansion, have worse estimate on the usefulness of terms, and also the performance could be more variant and less robust. If we have a large number for documents. We will start to add more irrelevant terms into the candidate set and waste large computation capacity.

Example Examination:

- Example1: More fbDocs works better.
  Original query: 751: Scrabble Players. Applying 50 documents on this query gives us the largest improvement. The MAP improves from 0.1685 to 0.1743. As we can see there are only two terms in this query. Therefore, it is very useful to add more terms to this query. Words like "game" is added. If we only use 10 documents, "game" is only weighted 0.0009, but when we use 50 documents, "game" weights 0.0013.
- Example2: Less fbDocs works better.
  Original query: 711:Train station security measures. When changing from 10 to 50, the relative importance of "measure: and "police" decreases. But importance of "rail" "station" increases. Words in two expansions are almost the same. But their relative importance have large influence on the result.

3 Experiment 3: The number of feedback terms

3.1 Experimental Results

	Indri BOW, Reference System (*given*)	**Query Expansion, Reference System Initial Results**					
		Feedback Terms					
		5 (Exp-3a)	**10 (Exp-3b)**	**20 (Exp-3c)**	**30 (Exp-3d)**	**40 (Exp-3e)**	**50 (Exp-3f)**
P@10	0.6800	0.7000	0.7600	0.7500	0.7200	0.7300	0.7300
P@20	0.6450	0.6150	0.6200	0.6450	0.6550	0.6450	0.6500
P@30	0.5600	0.5300	0.5467	0.5733	0.5800	0.5800	0.5900
MAP	0.1934	0.1729	0.1866	0.2023	0.2055	0.2083	0.2088
win:loss	N/A	Loss	Loss	Win	Win	Win	Win

3.2 Parameters

Indri:mu=1500

Indri:lambda=0.01

fb=Indri

fbDocs=30

fbMu=0

fbOrigWeight=0.5

3.3 Discussion

Overall Discussion:

It seems the more terms the expansion use, the better the performance of the result is. And the number of terms mainly help to improve P@20 and P@30 instead of P@10. Since more terms are used, it is more likely that they match more documents. Moreover, since there are more terms used in the query, scores of documents are evaluated in smaller granularity. Therefore, it could be easier to rank documents that are less related and have less matched terms. But in order to rank documents that already match lots of part of the query, more detailed and precise evidences are needed. In this case, adding more related but not so relevant or precise terms could harm the performance.

Example Examination

- Example1

Query 802 has a large improvement by add more terms. The MAP improves from 0.05 to 0.20. The original query is 802:Volcano eruptions global temperature. Words like gass, lava, ash are added. And these words are closely related to volcano eruption.

- Example2

Almost performs of all query increase when we increase the number of terms into the query. The MAP of query 826 starts to decrease from 30 to 50. The original query is 826:Florida Seminole Indians. As the number of terms increases, words like data, were, people, rain, book appear. These words are very common in many documents and similar to stop words. Their appearance add errors to the query and influence the performance.

4 Experiment 4: Original query vs. expanded query

4.1 Experimental Results

	Indri BOW, Reference System (*given*)	**Query Expansion, Reference System Initial Results**					
		fbOrigWeight					
		0.0 (Exp-4a)	**0.2 (Exp-4b)**	**0.4 (Exp-4c)**	**0.6 (Exp-4d)**	**0.8 (Exp-4e)**	**1.0 (Exp-4f)**
P@10	0.6800	0.6600	0.7000	0.7400	0.7500	0.7500	0.7000
P@20	0.6450	0.6300	0.6555	0.6555	0.6555	0.6400	0.6300
P@30	0.5600	0.5567	0.5833	0.5967	0.5733	0.5467	0.5300
MAP	0.1934	0.1971	0.2099	0.2121	0.2049	0.1920	0.1806
win:loss	N/A	Win	Win	Win	Win	Loss	Loss

4.2 Parameters

retrievalAlgorithm=Indri

Indri:mu=1500

Indri:lambda=0.01

fb=Indri

fbDocs=30

fbTerms=50

fbMu=0

4.3 Discussion

Overall Discussion

The larger fbOrigweight is, the more the query will rely on the original query. In our experiment, we have the highest performance when fbOrigweight is 0.4. It seems that the expanded query learned from the data fits in the information need more than the original query. But at the same time, the expanded query will usually contain terms from the original query.

Example Examination

Some queries are very sensitive to this parameter. Others are not. For example, query 758 is not so sensitive to this parameter. 758 : Embryonic stem cells. Its expansion also contains cell, stem, embryonic, which means the original query and the expansion are not so different. Therefore, the weight does not have much influence on the performance.

For some queries, the MAP score with weight 1 is higher than the score with weight 0, which means the quality of original query is better than the expansion (711,751,785,764,811,826). On the other hand, there are queries have a better expansion than the original one. But if we combine them together, we can have a higher score. For example, query 809 has highest score when weight is 0.4. The original query is 809:wetlands wastewater treatment. And there are words like “construct”, “quality”, “design” in the expansion. We can assume these two parts in the query matches different things and the weight combine them together.

5 Experiment 5: Indri vs. BM25 Query Expansion (11-742 Only)

5.1 Experimental Results

Present your experimental results in the format shown below. Report win:loss compared to no query expansion for Indri or BM25.

Your .zip / .tgz file must include files named HW3-Exp-5a.qry, HW3-Exp-5a.param, etc., in the QryEval directory. The experimental results must be reproducible by these files and the parameter values documented below.

	Indri			**BM25**		
	Reference System (*given*)	**(Exp-5a)**	**(Exp-5b)**	**Initial Retrieval (Exp-5c)**	**(Exp-5d)**	**(Exp-5e)**
fbdocs	N/A			N/A		
fbterms	N/A			N/A		
P@10						
P@20						
P@30						
MAP						
win:loss	N/A			N/A		

5.2 Parameters

Briefly document the values of other parameters used during this experiment.

5.3 Discussion

Analyze the experimental results.