Error Analysis For Sentence Scoring Component

Strategy for Identification:

Appearance of Correct Answer Token in Candidate Sentences(Filter out stop words)

(Further improvement: may consider synonym for biological terms)

Result Summary(12-test-alzheimer Doc)

detailed results please refer to QAresult_SentenceErrorAnalysis.txt

Sentence ErrorAnalysis for DocQA4MRE-2012_BIOMEDICAL_GS.xml_1: 0.06999999999999999

Sentence ErrorAnalysis for DocQA4MRE-2012_BIOMEDICAL_GS.xml_2: 0.08000000000000000

Sentence ErrorAnalysis for DocQA4MRE-2012_BIOMEDICAL_GS.xml_3: 0.10000000000000000

Sentence ErrorAnalysis for DocQA4MRE-2012_BIOMEDICAL_GS.xml_4: 0.09999999999999999

By observation of the statistics, the overall accuracy for selecting candidate sentences is pretty low. On average, for 10 candidate sentences, less than one of them is relevant to our correct answers. Of course, it may due to that we haven't considered synonym of answer terms.

Besides, even those sentences that are relevant to our correct answers are not ranked the highest.

In conclusion, we decide to present 2 solutions for this problem.

First,tune solr search engine for better returned candidate sentences if there is any in the document.

Second, expand the corpus in which we search for candidate sentences.

We decide to try the first solution first until we can gain no more better performance by tuning. Then we go to second solution and see if the strategies we used for first solution could be a general solution to large corpus. If not, we try to improve the strategies for better performance on larger corpus.

For second solution, we consider use Lucene for candidate document retrieval if there's large pool of candidate documents.

Solr Query Approach & Tuning: