1. Manual examination of TREC data and search results

From my examination of using various search queries consisting of either Titles, Descriptions or Narratives, I noticed that it is only the titles that appear to return relevant results which are similar to those in the “trec678.res” file. While the order was not exactly the same, the top 20 documents returned by my searches usually contained only documents in the top 40 - 50 ranked documents from the trec results file. In nearly all cases the top 10 ranked documents were included in the top 20 or 30 search results returned by my searches. It is my understanding that the exact ranking specified by the trec file cannot be achieved purely through standard searches, some understanding of the topic (such as a human reader would have) may be necessary to achieve the same level of accuracy.

2. Exploring evaluation metrics

The values for mean average precision are quite low at 0.2145, however this is understandable when you look at the precision at different rank cutoffs. At cutoffs of 5, 10 and 20 documents the precision is in the range of 0.4240 - 0.3487. This is pretty good considering a user is unlikely to look at more than 3 or 4 documents let alone 20. The precision at later documents numbers is much lower (less than 0.2) and this is what is skewing the mean average position. For recall, only about 50% of the relevant documents were returned on average. As mentioned in the previous section, this is likely due to a lack of context surrounding the search which would provide an understanding of the search topics, something which a human has that will allow more appropriate ranking of searches. There is a compromise between recall and precision here, which results in a slightly higher precision in the lower cutoffs (5 - 20) but a lower overall recall of relevant documents

Looking at results for individual topics gives proof to the above idea. Even looking at the searches 301 and 302 we can see this. For search 301, the topic of association and collaboration between crime organizations and countries is something that is difficult to describe textually, and may require a human to analyze and think about possible links in an abstract manner. It may also require a previous knowledge of an organization’s criminal activity which may not be mentioned in every document they appear in. This results in a low overall recall (only 74 of 474 documents) and a well below average mean average precision. For search topic 302, Poliomyelitis and Post-Polio, the rarity of these words and the fact that they are diseases means that any talk about them will be likely to be relevant. This makes finding the right documents a lot easier in this case as a word match is all that will be required. This can be seen in the recall results (64 out of 77 documents) and an above mean average precision of 0.4737. The precision at cutoffs is very high also, maintaining ranges of 0.6 - 0.7 within the top 30 documents returned. The precision does not fall far below average until after the 200 document cutoff point.