Assignment 2

Task 1.1

In automatic local analysis we only evaluate the document returned by the query. However the global analysis in done on the entire document collection. Both are automatic since the evalutaion is done without a user.

Task 1.2

Relevance feedback is about gathering user feedback on the given query. This feedback can be in one of three categories; explicit feedback, implicit feedback and pseudo feedback. This feedback is used to determine if the results are relevant for a new query.

Query Expansion: Is about expanding the query result based on the user input. This is done by looking up synonyms, spelling errors, semantically related words etc. By doing this the returned query is expanded with possible documents that might be useful to the user.

Term Re-weighting: This consept is about lowering the weight of terms on irrelevant documents, and increasing the weight of terms in relevant documents.

The main difference between query expansion and Term re-weighting is that term re-weighting is does not extend the returned query.

Task 2.1

Language model: This is a way to determine the best query for returning a know document. By having the probability of each term in a document, the language model, its possiable to generate the best possiable queries with highest probability of returning this document.

Task 2.2

d1 = An apple a day keeps the doctor away = 8
d2 = The best doctor is the one you run to and can’t find = 12
d3 = One rotten apple spoils the whole barrel = 7
Total words in collection = 27

q1 = doctor
q2 = apple orange
q3 = doctor apple

$$\hat{P}(t|Md) = (1{-}\lambda)\hat{p}_{mle}(t|Md) + \lambda\hat{p}_{mle}(t|C), \lambda = 0.5.$$

$$\hat{P}(q_1|d_1) = 0.5 * (\frac{1}{8} + \frac{2}{27}) = 0.099$$

$$\hat{P}(q_1|d_2) = 0.5 * (\frac{1}{12} + \frac{2}{27}) = 0.078$$

$$\hat{P}(q_1|d_3) = 0.5 * (0 + \frac{2}{27}) = 0.037$$

$$q_1 : d_1 > d_2 > d_3$$

$$\hat{P}(q_2|d_1) = 0.5 * (\frac{1}{8} + \frac{2}{27}) * 0.5 * (0 + 0) = 0$$

$$\hat{P}(q_2|d_2) = 0.5 * (0 + 0) * 0.5 * (0 + 0) = 0$$

$$\hat{P}(q_2|d_3) = 0.5 * (\frac{1}{7} + \frac{2}{27}) * 0.5 * (0 + 0) = 0$$

$$q_2 : d_1 = d_2 = d_3 = 0$$

$$\hat{P}(q_3|d_1) = 0.5 * (\frac{1}{8} + \frac{2}{27}) * 0.5 * (\frac{1}{8} + \frac{2}{27}) = 0.009$$

$$\hat{P}(q_3|d_2) = 0.5 * (\frac{1}{12} + \frac{2}{27}) * 0.5 * (0 + \frac{2}{27}) = 0.003$$

$$\hat{P}(q_3|d_3) = 0.5 * (0 + \frac{2}{27}) * 0.5 * (\frac{1}{7} + \frac{2}{27}) = 0.004$$

$$q_3 : d_1 > d_3 > d_2$$

Task 2.3

Smoothing is used to avoid zero values. This ensures that because one of the query terms have zero occurances in a document it automaticly does not make the entire score zero. However if there acctually is no occurances the score will become zero. This is shown in query 2 with orange never appearing in the collection. However as seen in query 3 document 2, apple is not in this document. However because we are using smoothing this probability score is not zero and gives a better overall result.

Task 3.1

Precision and recall is used to measure how well the IR system works. Where precision measures how many documents that are retrieved are relevant. Recall on the other hand measures the percentage of the relevant documents the system has found.

$$Precision = \frac{Relevant\ documents\ retrieved}{Retrieved\ documents} = P(relevant|retrieved)$$

$$Recall = \frac{Relevant\ documents\ retrieved}{Relevant\ documents} = P(retrieved|relevant)$$

Name	Relevant	Non-Relevant
Retrieved	TP (True Positive)	FP (False Positive)
Not Retrieved	FN (False Negative)	TN (True Negative)

$$Precision = P = \frac{TP}{TP + FP}$$

$$Recall = R = \frac{TP}{TP + FN}$$

Task 3.2

Mean Average Precision or MAP is the mean of the average precision score for each query. This gives an average precision per returned query and is can be translated to a single metric that represents the area under the Precision-Recall curve. MAP is also great at weighting errors happening higher up in the returned query and less weight for errors further down. Since MAP gives a single metric its really good for binary rating, however numeric rating is unachievable with this method of scoring a query. We lose a lot of information with MAP as well that might be useful for research.

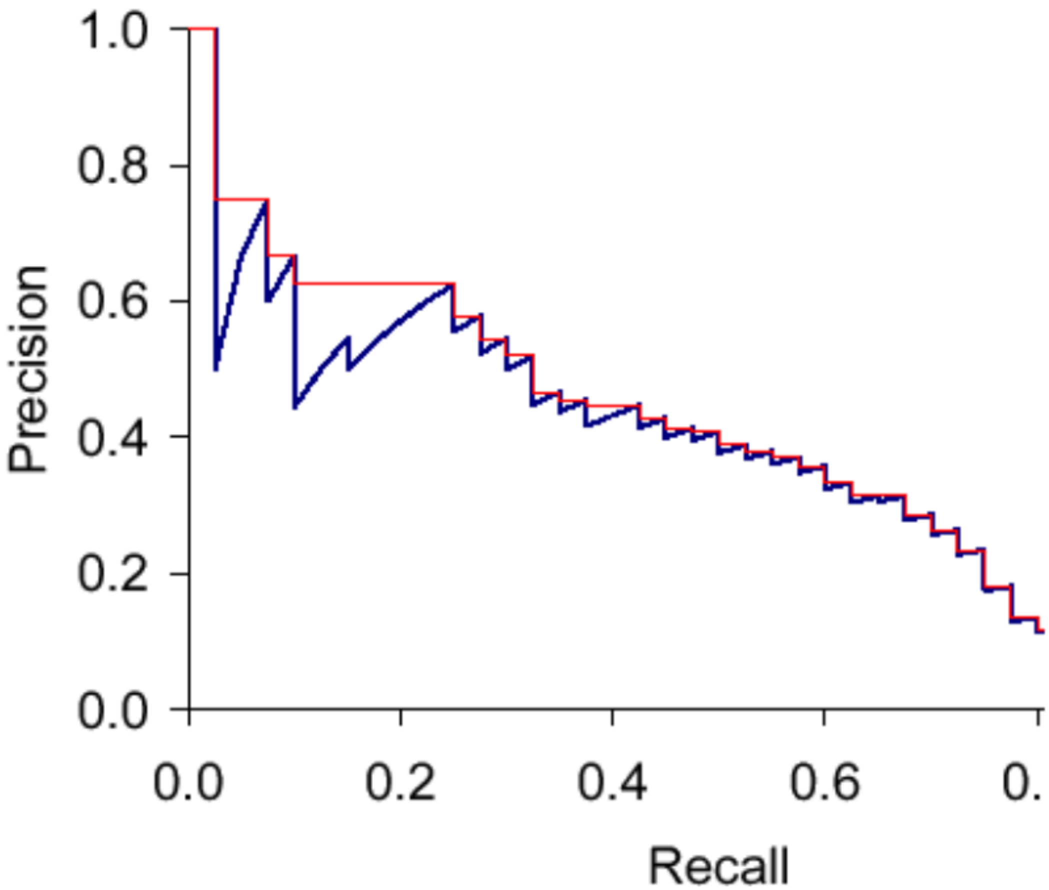
Mean Reciprocal Rank or MRR is a simple metric where is measures where the first relevant document is. This is a super simple method to compute and is a good method if you want to measure relevant items as lose to the top as possible. However this metric does not take into account the rest of the returned query. It also only relies on the first item hit, so a query with lots of hits and a query with only one hit will have the exact same score if both have the same first hit.

Task 3.3

Name	Relevant	Precision	Recall
55	☐		
500	☑	0.5	0.1
2	☐		
23	☑	0.5	0.2
72	☑	0.6	0.3
79	☐		
82	☑	0.57	0.4
215	☐		

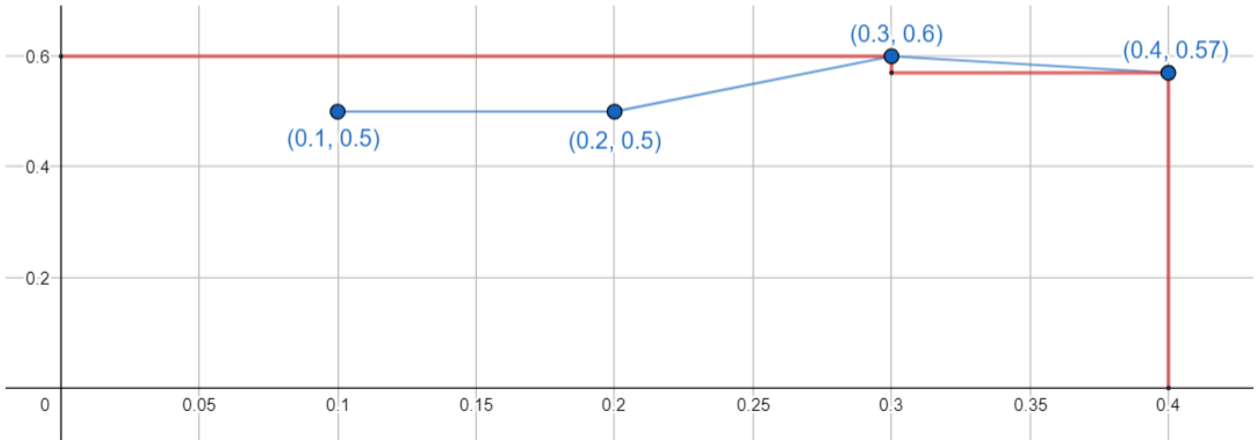
Task 4.1

Interpolated precision is used to smooth out the saw tooth shaped precision-recall graph. This is done by picking a recall level r and for all other recall levels r' we want to make sure that $P(r') >= P(r)$ where $P(r)$ is the precision of rank r. This ensures that the graph moves in a zigzag pattern insted of a saw tooth pattern and we get a lot smoother graphs.



As shown here the blue graph is the precision-recall graph with the saw tooth shape and the red graph (interpolated precision-recall graph) always cut off at the same precision level as the peak of the next recall level.

Task 4.2



Where the blue graph ish the precision-recall graph and the red dotted line is the interpolated graph. Here we can se that the graph skips the lower values and starts at 0.6 and then lowers.