Assignment 2

Language Model

1. For an information retrieval system we are looking to find what type of language models to use. Do you recommend using a Unigram language model or Higher order models? Provide justification for your recommendation.

2. Given the following documents and queries, build the language model according to the document collection

d1 = An apple a day keeps the doctor away.
d2 = The best doctor is the one you run to and can't find.
d3 = One rotten apple spoils the whole barrel.

q1 = doctor
q2 = doctor apple

Use MLE for estimating the unigram model and estimate the query generation probability using the Jelinek-Mercer smoothing

$$\widehat{P}(t|M_d) = (1-\lambda)\widehat{P}_{mle}(t|M_d) + \lambda\widehat{P}(t|M_c) \quad , \quad \lambda = \frac{1}{2}$$

For each query, rank the documents using the generated scores.

d_1 has 8 terms, d_2 has 12 terms, and d_3 has 7 terms. In total there are 27 terms.

$$simplified\ equation = \frac{\frac{tf_{d,t}}{|d|} + \frac{cf_t}{|T|}}{2} \quad as\ \lambda\ is\ \frac{1}{2}$$

	d_1	d_2	d_3
q_1	0.0995	0.0787	0.0370
q_3	0.0099	0.0029	0.0040

$$\widehat{P}(t|M_d) = \frac{tf_{d,t} + \mu P(t|M_c)}{|d| + \mu}$$

	d_1	d_2	d_3
q_1	0.0995	0.0796	0.0395
q_3	0.0099	0.0024	0.0042

3. Why did we use smoothing? Which query would have been zero if we did not use the smoothing?)

If we were not to use smoothing, the system would retrieve documents that don't contain all terms in the user query. Then, q1 would not return d2 at all due to this property, same with q2 which would return any documents. Although, if the term does not exist in the vocabulary at all the result still yields a value of zero.

4. When does Dirichlet Smoothing perform better? Apply the Dirichlet Smoothing with μ = 8 for all the queries.

When the document term lengths are smaller Dirichlet retrieves a higher score for the query, which could be useful for finding relevant documents for part of the query; whereas Jelinek-Mercer would score shorter documents as less relevant if it only matches part of the query.

Relevance Feedback

1. Explain the difference between automatic local analysis and automatic global analysis.

Local analysis analyses the local documents for query expansion, this is done by using the retrieved documents by a given query. Global analysis on the other hand uses the entire document collection for its analysis. Using global analysis the entire collection is used for building the thesaurus, whilst local does query expansion upon the search itself.

2. What is the purpose of relevance feedback? Explain the terms Query Expansion and Term Re-weighting. What separates the two?

Relevance feedback is used to readjust term weights to give the user a more relevant result to their query, in addition to query expansion. Query expansion is often done to extend the retrieved relevant documents, usually by expanding on the query by also searching for synonyms and semantically similar terms.

Evaluation of IR Systems

1. You are the judge in an information retrieval competition. Your task is to determine which team has a better performance. Two teams submit their results for the competition query Q as described in the paper below. In the ranking: R denotes a relevant document and N denotes a non-relevant document. In total there are 4 relevant document for the query Q. Compute the following evaluation measures for the ranking provided by Team1 and Team2:

a. *Precision at 6 (P@6) for Team1 and Team2, which team won?*
b. *Recall at 6 (P@6) for Team1 and Team2, which team won?*
c. *Mean Average Precision (MAP) for Team1 and Team2. Which team won?*
d. *R-Precision for Team1 and Team2. Which team won?*
e. *Mean Reciprocal Rank (MRR) for Team1 and Team2. Which team won?*

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Team 1	*N*	*R*	*N*	*N*	*R*	*R*	*R*	*N*	*N*	*N*
Team 2	*R*	*N*	*R*	*N*	*N*	*N*	*R*	*N*	*R*	*N*

	P@6	R@6	MAP	R-precision	MRR
Team 1	50%	75%	0.4929	25%	0.5000
Team 2	33%	50%	0.5099	50%	1

P@6: Team 1 =: $\frac{3}{6}$ *– Team 2 =* $\frac{2}{6}$ *– Recall: Team 1 =* $\frac{3}{4}$ *– Team 2 =* $\frac{2}{4}$

MAP: Team 1 = $\frac{\frac{1}{2}+\frac{2}{5}+\frac{3}{6}+\frac{4}{7}}{4} = 0.4929$ *– Team 2 =* $\frac{\frac{1}{1}+\frac{2}{3}+\frac{3}{7}+\frac{4}{9}}{4} = 0.5099$

R-precision: Team 1 = $\frac{1}{2}$ *– Team =* $\frac{1}{1}$ *as team 1's first relevant is doc2, and team 2 doc1*

Team one had both higher precision and recall, whilst Team 2 had higher MAP, R-precision and MRR

2. Describe the difference between how differently Precision and Recall can evaluate the retrieval quality of an IR system.

Precision simply tells us how many of the currently displayed documents are relevant or not; recall on the other hand tells us if the relevant documents to our query have been retrieved, regardless of where they are in the order displayed. Recall will thus give us a more precise measurement of whether the system has managed to retrieve the relevant documents, rather than if they have been ordered correctly.

Interpolated Precision

1. What is interpolated precision?

Interpolated precision reflects the maximum precision corresponding to the recall level. It shows the maximum precision up until it is actually reached by the recall, i.e. if recall is 1 of *n* at the *j-th* position, precision is the same up until this level at *j - 1, j - 2* and so on until the first retrieved document.

2. Given the following set of relevant documents

rel = {23, 10, 33, 500, 70, 59, 82, 47, 72, 9},

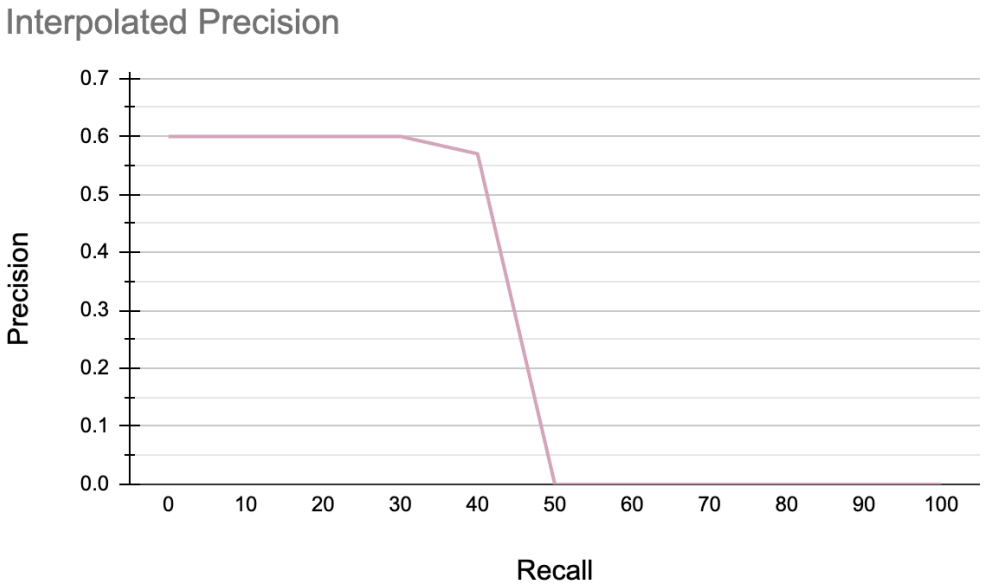
and the set of retrieved documents

ret = {55, 500, 2, 23, 72, 79, 82, 215},

find the interpolated precision and make a graph.

Position	Document	Position	Document
1	55	5	72
2	500	6	79
3	2	7	82
4	23	8	215

***P2**: P = 0.5, R = 0.1 – **P4**: P = 0.5, R = 0.2 – **P5**: P = 0.6, R = 0.3 – **P7**: P = 0.57, R= 0.4*



1. The figure below depicts interpolated precision-recall curves for two search engines that index research articles. There is no difference between the engines except in how they score documents. Imagine you're a scientist looking for all published work on some topic. You don't want to miss any citations. Which engine would you prefer and why?

Engine 2 would be the most preferable to not miss any citations as this engine's interpolated precision does not fall off to zero; leading me to believe it does retrieve all the relevant documents on the contrary to Engine 1.