Assignment 03
Digital Libraries and Foundations of Information Retrieval
Winter semester 2022

1542011 Franka Brunen, 1365848 Andreas Schneider

Task 1:   Information Needs   3+6+6 Points

(a)   1 Information on the person „Queen Elizabeth"

         2 Information on the term „queen", as in „queen regnant"

         3 Information on the term „queen", as in the music band

         4 Information on the term „queen", as in the magazine

         5 A disambiguation including all of the above and more like `https://en.wikipedia.org/wiki/Queen`

(b) First 10 results for duckduckgo.com:

Name	1	2	3	4	5	other
Queen (Band) - Wikipedia			X			
QueenOnline.com - The Official Queen Website			X			
Queen - Greatest Live Performances - YouTube			X			
Queen - Bohemian Rhapsody (Official Video Remastered) - YouTube			X			
Queen — Biografie - Universal Music			X			
Elisabeth II. - Wikipedia	X					
Queen Elizabeth (†): Enthüllt! Ihre strikte Regenschirm-Regel	X					
Beisetzung von Queen Elizabeth II.: ++ Sarg im Familienkreis beigesetzt ...	X					
Queen-Beerdigung: Der Tag in Bildern - ZDFheute	X					
Queen - Wikipedia, la enciclopedia libre			X			

First 10 results for duckduckgo.com:

Name	1	2	3	4	5	other
Queen (band) - Wikipedia			X			
Elizabeth II - Wikipedia	X					
Queen The Band Now Worth More Than The Queen Of England - Audacy	X		X			
Adam Lambert confesses fear of following Freddie Mercury before ...			X			
Queen + Adam Lambert - Wikipedia			X			
Behind the Band Name: Queen - American Songwriter			X			
Queen Official - YouTube			X			
QueenOnline.com - The Official Queen Website			X			
Queen — Spotify			X			
Queen - Facebook			X			

Aggregation of results:

	1	2	3	4	5	other
duckduckgo.com	4	0	6	0	0	0
google.com	2	0	9	0	0	0

In one case, row 3 from google.com, an article could be assigned to two information needs: **1** and **3**.

(c)   1 „queen" + „ elizabeth"
         → 10 0 0 0 0 0

         2 „queen" + „ regnant"
         → 1 5 0 0 0 4
         (with two „other" being info on queens other than Elizabeth II and two being lists of queen regnants)

3 „queen" + „ band"
→ 0 0 10 0 0 4

4 „queen" + „ magazine"
→ 2 0 3 4 0 1
(with „other" being „GUIDOS DEKO QUEEN - das Magazin für alle Deko-Fans";
the Term „magazine" is obviously not very effective in discerning the old UK Queen
Magazine)

5 „queen" + „ disambiguation"
→ 0 0 1 0 3 6
(with all „other" being disambiguation wiki pages of fictional universes)

Task 2:   Precision and Recall   5+5+5 Points

(a) **Information need:** The difference between \tabular and \tabular*
Search query: „tabular vs tabular*"

(b) **Precision:** $1/10 = 0.1 = 10\%$

(c) One could say that the data to be searched is infinite, as new web pages are indexed at every
moment. Therefore the amount of false negatives is unknown.

Between two or more search engines, one can accumulate all the true positives and compare
the amount of false negatives of each search engine. If one true positive is returned by one
search engine but not the other, the other has a false negative.

Task 3:   Additional evaluation metrics   4+6+5 Points

(a) Definitions:

- F_β „is a measure of a test's accuracy. It is calculated from the precision and recall of the test" [1]

- *Fallout* is „the fraction of non-relevant documents that are retrieved $\frac{|ret \cap \neg rel|}{|\neg rel|}$" [2]

The impact of the parameter β is that it defines the weight of recall in relation to precision. In other words, recall is weighed „β times as much" as precision.

Explanations:

- $F_\beta = 0$ is possible and means that no relevant document was retrieved.

- $F_\beta = 1$ is possible and means that no non-relevant document was retrieved and either β or the number of non-retrieved relevant documents is 0.

- $Fallout = 0$ is possible and means no non-relevant documents were retrieved.

- $Fallout = 1$ is possible and means all non-relevant documents were retrieved.

(b) Determinations:

Result set	Recall	Precision	Fallout	F_1
S_1	4/9	0.40 = 4/10	5/21	0.53 = 10/19
S_2	7/9	0.54 = 7/13	6/21	0.64 = 14/22
S_3	6/9	0.75 = 6/8	2/21	0.71 = 12/17

In terms of these metrics, S_3 is the best-performing one. With this specific search query and document collection, the IR retrieved most relevant results and few non-relevant ones. Only in *Recall* is S_2 better, but not enough to compensate for the false positive and -negatives that affect its F_β.

(c) Assuming that there is always 1 relevant and 1 non-relevant document to any search query:

- An IR system that retrieves all documents on every query would always have a recall of 1, as it would always retrieve all relevant documents.
  Its precision would always be $\frac{rel}{|D|}$; relevant documents divided by the amount of all documents, as it would always return all relevant documents alongside all others.
  Its fallout would always be 1, as all documents contain all non-relevant documents.

- An IR system that retrieves no documents on every query would always have a fallout of 0, as it would never retrieve any non-relevant document.
  Its recall would always be 0, as it would always return 0 relevant documents.
  Its precision would always be 0, as it would always return 0 relevant documents.

---

[1] Wikipedia `https://en.wikipedia.org/wiki/F-score`

[2] L. Egghe (2008) The measures precision, recall, fallout and miss as a function of the number of retrieved documents and their mutual interrelations `https://www.sciencedirect.com/science/article/pii/S0306457307001598`