Assignment 8
Digital Libraries and Foundations of Information Retrieval
Winter semester 2022

1542011 Franka Brunen, 1365848 Andreas Schneider

Task 1:   Relevance Assessment   2+4+2+7 Points

(a,b)

	(a)	(b)
ID	**Title**	**Relevance**
1	University_of_Trier	highly relevant
2	Trier	relevant
3	Lars_von_Trier	not relevant
4	Trier_University_of_Applied_Sciences	relevant
5	List_of_universities_in_Germany	relevant
6	History_of_Trier	relevant
7	New_Trier_High_School	not relevant
8	Antichrist_(film)	not relevant
9	Jost_Trier	not relevant
10	Electorate_of_Trier	relevant
11	List_of_New_Trier_High_School_alumni	not relevant
12	Roman_Bridge_(Trier)	not relevant
13	Trier_witch_trials	relevant
14	ECB	not relevant
15	Trier_Cathedral_Treasury	not relevant
16	Seamless_robe_of_Jesus	not relevant
17	Herman_Trier	not relevant
18	Porta_Nigra	not relevant
19	Walter_Trier	not relevant
20	Lamprey	not relevant

(c) Precision of the above query result set: 7/20

(d) `<title>` *norden*
`<desc>` Anything about the town called Norden in Germany
`<narr>` Highly relevant documents must be primarily about anything concerning the town Norden in Germany. Relevant documents must at least reference the distinct town.

ID	Title	Relevance
1	Norden	relevant
2	Carl_Norden	not relevant
3	Norden_bombsight	not relevant
4	Denis_Norden	not relevant
5	S_Norden	not relevant
6	It'll_Be_Alright_on_the_Night	not relevant
7	Norden_Systems	not relevant
8	Foreningen_Norden	not relevant
9	Coop_Norden	not relevant
10	Peter_Van_Norden	not relevant
11	Tommy_Norden	not relevant
12	Norden_(surname)	not relevant
13	Frederic_Louis_Norden	not relevant
14	John_Barrymore	not relevant
15	Eduard_Norden	not relevant
16	Norden,_Lower_Saxony	highly relevant
17	Christine_Norden	not relevant
18	Norden,_Nebraska	relevant
19	PostNord	not relevant
20	Putnam_model	not relevant

Precision of the above query result set: 3/20

Task 2: Relevance Assessment 1+2+3+3+3+3 Points

(a) The relevance of a document for a topic is not unambiguously assessable.

(b) To solve this, the first n results of multiple search engines can be combined. After removing duplicates, assessors can be tasked to assess documents from the pool in a random order.

If this question pertains to documents not contained in a fully assessed pool, one can say that they can be ignored, as all participating search engines will have ranked them less relevant than its preceding n documents.
If this question pertains to documents inside a not fully assessed pool, I don't quite understand what is asked. One could say that every assessment is just as likely to be about the most relevant document as the other. So if half the documents were assessed, there is a 50% chance that „the most relevant document" is already among the assessed documents. Also all documents, assessed or not, have been deemed among the n most relevant documents by the participating search engines.

(c) If the new search engine produces results that vary greatly from the search engines underlying the existing collection, it could seem to have bad quality.

One could assess documents that are new arrivals without any bias and only then judge the new search engine and re-judge the old search engines' quality.

(d) Without further details, one cannot declare either 10 topics with 1000 results each or 100 topics with 100 results each „better" for a benchmark for text search.

If the retrieval task is recall-oriented, assessing more results on fewer topics becomes more attractive. If the retrieval task is precision-oriented, assessing fewer results on more topics is preferable.

If one of the selected systems changes its behavior after n documents and, for example, then returns results to only part of the query, or anything but documents sorted by the same metric as the first n documents, assessing more topics with fewer results each, according to the irregular behavior, will produce more homogeneous and comparable assessments.

(e) If one assumes that the user reads the results from top to bottom, is able to recognize a relevant document and clicks on the first one they see, it could appear as if the second result

was the first relevant document.
Were that the case, the first result could still be relevant if, for example, the user was unable to recognize relevant documents, or they prioritized and clicked on the second result, because they were successfully clickbaited by a non-relevant result, or if they do not read the results from top to bottom, or if they do not click the first result that seems relevant, or one of many other reasons.

(f) If one assumes that the user can recognize relevant results and will click them, it could appear as if the third result was not relevant.
Were that the case, the third result could still be relevant, if the user did not click at all. Even if they clicked other results, and even if results before and after the third result were clicked, one could assume that the user clicked randomly, did not recognize the third result, or found a satisfactory solution, before they would click on the third result.

Task 3: MAP 6+6+3 Points

(a)

A	V	R	**C**	S	**B**	J	L	E	**D**
$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{4}$	$\frac{2}{5}$	$\frac{3}{6}$	$\frac{3}{7}$	$\frac{3}{8}$	$\frac{3}{9}$	$\frac{4}{10}$

$AP_{q_1} = (\frac{1}{1} + \frac{2}{4} + \frac{3}{6} + \frac{4}{10})/4 = \frac{24}{10}/4 = \frac{24}{40} = \frac{6}{10} = 0.6$

(b)

A	V	**R**	C	**S**	B	J	L	E	D
$\frac{1}{1}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	$\frac{3}{5}$	$\frac{3}{6}$	$\frac{3}{7}$	$\frac{3}{8}$	$\frac{3}{9}$	$\frac{3}{10}$

$AP_{q_2} = (\frac{1}{1} + \frac{2}{3} + \frac{3}{5} + 0)/4 = \frac{34}{15}/4 = \frac{34}{60} = \frac{17}{30} \approx 0.57$

(c) $MAP = (\frac{6}{10} + \frac{17}{30})/2 = \frac{35}{30}/2 = \frac{35}{60} = \frac{7}{12} \approx 0.58$

$GMAP = \sqrt[2]{\prod_{i=0}^{2} q_i} = \sqrt[2]{\frac{6}{10} * \frac{17}{30}} = \sqrt[2]{\frac{102}{300}} = \sqrt[2]{\frac{28}{75}} \approx \frac{5.29}{8.66} \approx 0.61$