Task Sheet 2
Evaluation of IR Systems and NLP

1. Let (0.75, 0.25) be the recall-precision tuple for a query q. These values occurred after 3 relevant documents were found in the ranking order of the result list.
   a) Determine the amount of the documents that were observed until this point!
   b) Determine the total amount of relevant documents in the collection!
2. Explain the *F-Measure*!
3. Consider two IR systems that are being compared using a query. There are exactly 10 relevant documents for this query and only the first 15 result documents are considered. The result is a ranking, where a + denotes a relevant document and a − an irrelevant one:
   - System 1: + − + − − − + + + − − − − + +
   - System 2: − − − + − − − + − + − + + + −

   This means, that the first document in the result ranking of system 1 is relevant, the second one is irrelevant, the third relevant, and so on.

   Determine the recall and precision values of each of the systems after for every time a relevant document was found (P@0.1, P@0.2, ...)! Then, determine the average precision.
4. Assume that the total amount of relevant documents in the last task is now not 10 but 15. How would the values P@0.1, P@0.2, ... be determined now?
5. Explain what is meant with *Mean Average Precision*!
6. What is the *Discounted Cumulative Gain* and what is it used for?
7. Consider the following gradual relevance assessment by a user. The values are in the range [0,4], with 0 meaning irrelevant. Calculate DCG_p, $IDCG_p$ and $nDCG_p$ for $p = 6$!

Rang	1	2	3	4	5	6
Relevanz	2	1	0	4	3	0

8. Briefly explain what is meant by a *stop word list* and how it affects precision and recall of an IR system! How would you implement it for an IR system?
9. Are there alternative methods to the stop word list?
10. What is meant by *inflection* and why does it pose a hurdle to information retrieval?