HOMEWORK 4:

RANKING MODELS AND EVALUATION

HOMEWORK 4 (TASK 1):

RANKING MODELS

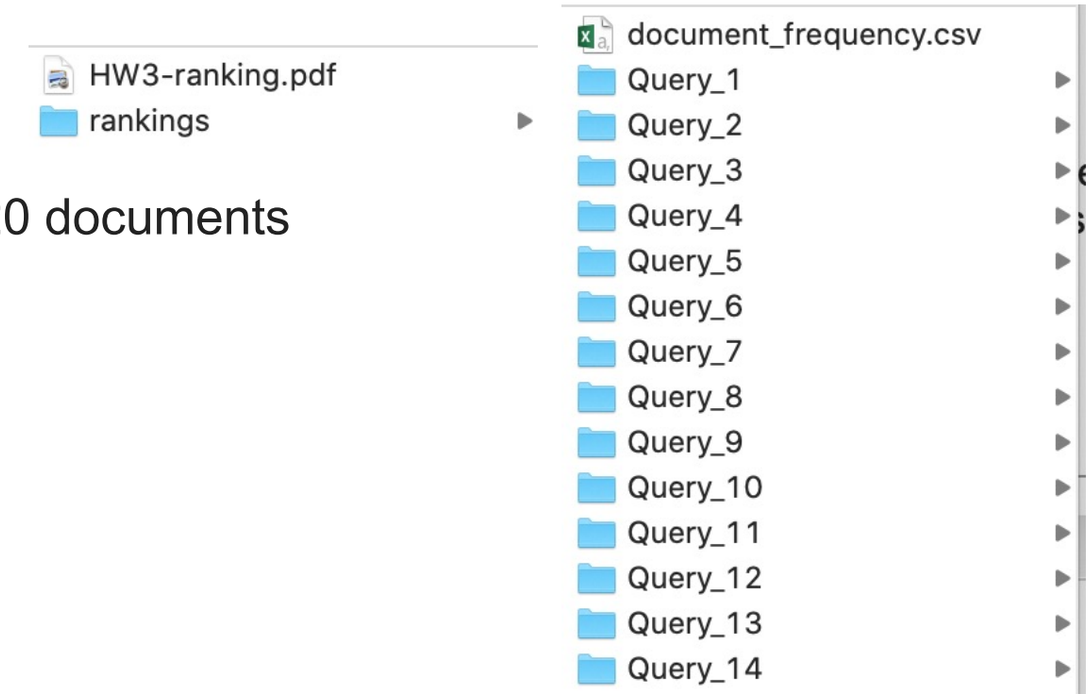
KEY - OVERVIEW

- Develop application to rank crawled data with Vector Space Model and Okapi BM25
- Rank documents (21 queries results crawled in Homework 1)

DETAILS

STEPS

- In folder “rankings”, there are 21 folders “Query_num”.
- For each query, “urls.tsv” containing following fields for 20 documents
  - query text
  - query description
  - url
  - docID – use this to find the html file in folder “Documents” (-1 means the file is not found)



- Steps
  - Parse the html file to get the main body of the document. (Pure text, without html tag.)
  - Scan the query and the document to get *tf, qtf, dl, etc.* (Note that there is no collection for your homework, so we provide a “***document_frequency.csv***” for ***df*** of terms in queries. And we set ***N = 100 billion***, ***avdl = 500*** for the collection.)
  - Computing VSM score and BM25 score for each document.

SUBMISSION FILES IN TASK 1

- File name convention:

  RM_RankModelName_QueryNumber_StudentID.csv

  e.g. RM_BM25_1_20198888888.csv

  Rank model names options: VSM and BM

  QueryNumber: 1~21

  2 * 21 = 42 files

  - Each line in file contains following fields:
    - “query text”, “query description”, “url”, “docID” – just keep it as in the ***original files***
    - rank – Number (start from 1) – rank of the document according to ranking models
    - rankScore - Number (calculated rank score according to ranking models)

- Source code (C++, C#, Java, JavaScript, Perl, Python, PHP, Scala, ……)
  - Details of VSM & BM25

HOMEWORK 4 (TASK 2):

EVALUATION

KEY - OVERVIEW

- Calculate annotation consistency
- Evaluate performances of Search Engines and your ranking models (VSM, BM25) with MAP and NDCG
- Analyze the results in your report

DETAILS

KAPPA

Cohen's Kappa coefficient measure for inter-judge (dis)agreement

- Each query was annotated by 2/3/4 times. Please find the average of pairwise Kappa.
  - If there are 2 annotators A and B, compute Kappa = Kappa(A, B)
  - If there are 3 annotators A, B and C, compute k_1 = Kappa(A,B) k_2 = Kappa(B,C) k_3 = Kappa (A,C). Then the final result Kappa = mean(k_1, k_2, k_3)
  - If there are 4 annotators A, B,C and D, compute k_1 = Kappa(A,B) k_2 = Kappa(B,C) k_3 = Kappa (C,D) …. Then the final result Kappa = mean(k_1, k_2, k_3…k_6)

- You need to compute 2 Kappa scores:
  - 5-levels Kappa - you take into consideration annotation result as it is; -1,0,1,2,3
  - 2-levels Kappa - you assume -1,0 as irrelevant; 1,2,3 as relevant

MAP

MAP(11-point): Mean 11-point Average Precision

- Relevant document: the average of relevance annotation score >= 1
- Irrelevant document: the average of relevance annotation score < 1
- Set of relevant documents is union of relevant documents from Baidu results and Bing results. 1 (*if the query is annotated by more than 3 students, you can do this by voting)

NDCG

Normalized Discounted Cumulative Gain

- Document relevant score mapping:
  - R_m is the average of relevance annotation score
  - $R_m >= 2$ map to 3
  - $0.5 < R_m < 2$ map to 2
  - $0 < R_m <= 0.5$ map to 1
  - otherwise map to 0

- Compute NDCG@5, NDCG@10

DETAILS INPUT FOR EACH QUERY

- ”relevance_annotation.csv”
- Each line in the file contains following fields:
  - query
  - url
  - docID
  - SE
  - Annotations ([annotation1; annotation2; …])

example

query	url	docID	SE	annotations
Coronavirus disease outbreak	https://www.who.int/emergencies/diseases/novel-coronav	61	BING	[1;3]
Coronavirus disease outbreak	https://www.cdc.gov/coronavirus/2019-ncov/index.html	62	BING	[1;3]
Coronavirus disease outbreak	https://www.cdc.gov/media/dpk/diseases-and-conditions	63	BING	[1;3]

DETAILS INPUT FOR EACH QUERY

- ”SE_ranking.csv”
- Each line in the file contains following fields:
  - rank
  - SE1_docID
  - SE2_docID

- “RM_ranking”
- The results you get in your HW4 TASK 1
- Useful Information
  - docID
  - VSM_rank
  - BM25_rank

example

rank	baidu_docID	bing_docID
1	1	11
2	2	12
3	3	13
4	4	14
5	5	15
6	6	16
7	7	17
8	8	18
9	9	19
10	10	20

SUBMISSION IN TASK 2

- File name convention:

  EVA_“query_number”_“student_id”.csv

  EVA_1_2016999999.csv

  - The file should contain following fields:
    - kappa2 - consistency score for 2-levels kappa
    - kappa5 - consistency score for 5-levels kappa
    - MAP - evaluation score for MAP@10 (2 SEs, 2 models)
    - NDCG@5 - evaluation score for NDCG@5 (2 SEs, 2 models)
    - NDCG@10 - evaluation score for NDCG@10 (2 SEs, 2 models)
    - 21 files

example

measure	score
kappa2	
kappa5	
MAP-baidu	
MAP-bing	
MAP-VSM	
MAP-BM25	
NDCG@5-baidu	
NDCG@5-bing	
NDCG@5-VSM	
NDCG@5-BM25	
NDCG@10-baidu	
NDCG@10-bing	
NDCG@10-VSM	
NDCG@10-BM25	

SUBMISSION: REPORT IN TASK 2

- File name convention:

  EVAREPORT_“student_id”.pdf

  - Annotation consistency
    - Compare your annotation with others
    - Analyze the annotation difference of the query you defined with others, and discuss the annotation difference between you individual and the third-party annotators.
  - SE performance,
  - Your Ranking model performance,
  - Evaluation difficulties and other details.

- Source code (C++, C#, Java, JavaScript, Perl, Python, PHP, Scala, ……)

SUBMISSION

- Two folders:
  - Folder1: Ranking_models
  - Folder2: Evaluation

  Compress them into one ZIP file named by “your_student_id”.zip

DEADLINE:

10:00am (UTC+08:00, Beijing Time) Apr. 5th, Monday
Submit your homework to web learning platform of thu:

Http://learn.tsinghua.edu.cn our course section “Assignment”(课程作业)

QUESTIONS ?