COMPSCI546_132815_FA20-Applied Information Retrieval Fall 2020

Programming assignment: Inference Network

The purpose of this project is to explore evidence combination on a small collection of documents. You will use the collection, your index, and your retrieval APIs from the previous project.

Combining Evidence

Implement the following belief functions as scoring functions within your retrieval API

- not
- or
- and
- weighted and
- sum
- weighted sum
- max

Implement the following structured query operators within your retrieval API

- ordered window
- unordered window
- filter require
- filter reject
- boolean and

Evaluation

We provide a number of "queries" specified in plain English. Please run these queries using the two phrase operators, ordered window and unordered window. For ordered, use a distance of 1 (exact phrase), for unordered, use a window width 3*|Q| (three times the length of the query). Please run these queries with each of the operators: SUM, AND, OR, and MAX. Use dirichlet smoothing with μ=1500 for all runs.

Please output the results of these queries in ***trecrun*** format as from the previous programming assignment, one output file for each different query operator.

Queries

- ***Q1***: the king queen royalty
- ***Q2***: servant guard soldier
- ***Q3***: hope dream sleep
- ***Q4***: ghost spirit

- ***Q5***: fool jester player
- ***Q6***: to be or not to be
- **Q7**: alas
- **Q8**: alas poor
- **Q9**: alas poor yorick
- **Q10**: antony strumpet

Grading Rubric

(5%) Submission is in the correct format.

A ***single*** archive file (zip, tar.gz) was submitted to Moodle and it contains at least the following contents:

- ***report.pdf*** - your report (see below)
- ***src/**** - your source code
- ***README***
  - It has instructions for downloading dependencies the code.
  - It has instructions for building the code.
  - It has instructions for running the code.
- ***od1.trecrun, uw.trecrun, sum.trecrun, and.trecrun, or.trecrun, max.trecrun***
  - Query Output Files: these should be in the trecrun format above.
  - Please generate a separate file for each run.
  - Each file should contain all 10 queries.

(50%) Source code that implements the evidence combination and structured operators.

Please be aware that you may lose points for problems in your code even if it appears to run correctly. We expect to see code used for all parts of the assignment in your submission.

(45%) The report discusses your work.

- (5%) Description of the system, design tradeoffs, questions you had and how you resolved them, etc. List the software libraries you used, and for what purpose.
- (10%) Look at the top ten results for ***Q6-Q10*** using the output from the ordered and unordered window operators**.** Using the same relevance judgment process as the previous retrieval models assignment, evaluate the difference between the two window operators. Explain any differences in behavior. Anecdotal evidence is sufficient.
- (10%) What needs to be done to implement a structured query language for your retrieval system. Sketch out the design.
- (20%) How do you expect each of the combination functions to behave, both individually, and in concert? What impact would normalization have on the **and** and **weighted and** operators?

Submission status

Submission status	No attempt
Grading status	Not graded
Due date	Friday, October 23, 2020, 11:59 PM
Time remaining	3 days 12 hours
Last modified	-

Submission comments

Comments (0)

Add submission

You have not made a submission yet

◀ Midterm Exam

Jump to...

Quiz: Evidence Combination ▶

You are logged in as Aarshee Mishra (Log out)
COMPSCI546 132815 FA20