Information Retrieval & Social Web

CS 525/DS 595

Worcester Polytechnic Institute

Department of Computer Science

Instructor: Prof. Kyumin Lee

Midterm

- The exam will be held in the next class.

- The exam is closed book.

- You may bring one standard 8.5" by 11" piece of paper with any notes you think appropriate or significant (front and back).

- No electronic devices allowed.

Previous Class…

Boolean Retrieval Model

Previous Class…

Boolean Retrieval Model

Inverted index

Previous Class…

Query Optimization

Previous Class…

Query Optimization

Preprocessing Documents
➜Tokenization, Normalization, Stemming, Stop words

Previous Class…

Query Optimization

Preprocessing Documents ➔Tokenization, Normalization, Stemming, Stop words

Skip pointers & Positional index

Previous Class...

Wild-card queries
➔ Permuterm Index

Previous Class...

TF and IDF

Previous Class…

TF and IDF

tf-idf weighting

$$w_{t,d} = (1 + \log \text{tf}_{t,d}) \cdot \log \frac{N}{\text{df}_t}$$

Previous Class…

Cosine Similarity

Previous Class…

Cosine Similarity

Computing scores in a complete search system

Previous Class…

Statistical Language Models

Crawler

Previous Class...

Statistical Language Models

Crawler

Web APIs

Previous Class...

PageRank

Evaluation