



Information Retrieval (CS317)

Instructor : Maryam Bashir

Credit Hours : 3, Class : BSCS Fall - 2016

Course Description:

This course is designed to help you to understand search engines, evaluate and compare them, and modify them for specific applications. It provides broad coverage of the important issues in information retrieval and search engines. Mathematical experience including basic probability is strongly desirable.

CourseText Book

1. [CMS] Search Engines by Croft, Metzler, and Strohman.
2. [MRS] Introduction to Information Retrieval by Manning, Raghavan, and Schütze - available free online.

Course Outline:

Topics
Introduction Key problems, Information need, Queries and documents, Matching scores, Inverted Index , Stopping, stemming
Indexing / Index Construction / Storage SPIMI and BSBI Algorithms for sorting index , Skipgrams, Co-occurrence, Index Storage, Distributed Indexes
Retrieval Models / Vector Space Vector-space model, Term weighting, Cosine measure, Tf-Idf, BM25, Co-occurrence, bigrams
Retrieval Models / Language Models Language generative Models, Query Likelihood, Model Divergence, Smoothing, Relevance Feedback
IR Evaluation/ Measures IR ranking performance, Set measures: Precision, Recall, F1, Accuracy, ROC, confusion matrix, Ranking measures: R-prec, AP, nDCG, Reciprocal Rank

Compression Delta , Omega, Elias Omega, Lempel Ziv, Huffman Codes, Index Compression
Text Statistics Zipf's Law, Heap's Law
Crawling Crawling Basics, HTTP Links, Graph BFS recap, Frontier/Queue , Duplicates
Web Retrieval Link analysis, Markov Chains, PageRank
Machine Learning / Features Document understanding, Features, Extracting Query Features, Similarity, How to measure ML, ML algorithms
Classification, Collaborative Filtering
Clustering Introduction to the Clustering, K-means clustering, Hierarchical clustering

Grading Policy:

The tentative distribution of marks is as under:

QUIZZES 5

PROGRAMMING ASSIGNMENTS 4

HOMEWORKS 6

PRESENTATION 5

MIDTERMS 30

FINAL 50

Plagiarism Policy

You are not allowed to copy code for programming assignments from internet or any other student. Penalty of plagiarism in programming assignments will be from one of the following depending on severity of case:

1. -1 absolute from final grade
2. Final grade is lowered
3. F in course

The weight age of all programming assignments is only 4 so if you do not want to spend time on it then best policy is not to submit it. Programming assignments are only for those students who want to learn the course material by applying it to practice. If you are only concerned about your grade then do not attempt these assignments, submitting plagiarized material for getting 4 marks is not worth the risk of plagiarism penalties.