

National University of Computer and Emerging Sciences
Information Retrieval (Undergraduate)
Spring 2014

Instructors: Awais Athar

Objectives: This course aims to develop an understanding of the data structures and algorithms required to produce document retrieval systems. At the end of this course, students should be able to understand main concepts of the major information retrieval approaches as well as the challenges in the tasks of web search, clustering, text classification.

Textbook and Reference Books:

- *Introduction to Information Retrieval*. C. Manning, P. Raghavan, and H. Schütze.
- *Modern Information Retrieval*, R. Baeza-Yates and B. Ribeiro-Neto.
- *Managing Gigabytes*. Ian Witten, Alistair Moffat, and Timothy Bell. Morgan Kaufman.
- *Foundations of Statistical Natural Language Processing*. Christopher D. Manning and Hinrich Schütze.

Tentative Lecture Plan:

Lectures	Topics	Assessments
1-2	Introduction, Key problems, Information need, Query, Inverted Indexing	Assignment 1: Build a search engine
3-4	Boolean retrieval, Tokenisation, Character encoding, Stemming, Lemmatisation, Stop words	
5-6	Vector-space model, Term weighting, Cosine measure, Weighting variants	Quiz 1
7-8	Corpus related problems, Creating test collections, Crawling, Crawler architecture	Assignment 2: Build a crawler
9-10	Kappa measure, Inter/intra-judge agreement, evaluation, User happiness, Precision, Recall, F-measure	Quiz 2
Mid 1		
11-12	Link analysis, Markov Chains, PageRank, Hubs and Authorities	Assignment 3: Detect spam mail
13-14	Introduction to text classification, Naive Bayes models, Feature Selection	Quiz 3
15-16	Vector space classification, Rocchio classification, k nearest neighbour	
17-18	Support vector machines, Kernel function, Evaluation of classification, Micro/macro-averaging	Quiz 4
Mid 2		
19-20	Introduction to the Clustering, K-means clustering, Hierarchical clustering	Assignment 4: Cluster newsgroups
21-22	Topic Extraction, Latent semantic Indexing	Quiz 5
23-24	Information Extraction, Sentiment Analysis	Assignment 5: Predict sentiment of tweets about a given topic
25-26	Question Answering, Summarisation	
27-28	Current trends in IR, course overview, project presentations	
Final Exam		

Evaluation

Quizzes	15
Assignments	25
Mid Term Exam	20
Final Exam	40
Total:	100 points