|  |  |
| --- | --- |
| Logo Fast | **NATIONAL UNIVERSITY**  **of Computer & Emerging Sciences, Lahore** |

Department of Computer Science

**CS317 – Information Retrieval**

**FALL 2018**

**Instructor Name:** Maryam Bashir

**Email address:** maryam.bashir@nu.edu.pk

**Office Location/Number:** C142

**Office Hours:** Wed, Fri 12:30 to 2:00 PM **,** Fri 10:00 am till 11:00 am

**Course Information**

**Program:** BS **Credit Hours:** 3 **Type:** Elective

**Pre-requisites (if any):** Data Structures, Probability and Statistics

**Class Meeting Time:** Wed, Friday: Section **A** 2:00 PM to 3:20 PM, Section **B** 11:00 AM to 12:20PM

**Class Venue:** Wednesday CS-1, Friday Section A CS-5 Section B CS-1

**Course Description:**

This course provides broad coverage of the important issues in information retrieval. It is designed to help you to understand how search engines work, how to build your own search engine, evaluate its performance, and modify it for specific applications. A number of advanced topics will be covered to address more recent developments in IR such as collaborative filtering and Topic Modeling. Students will furthermore acquire practical experience in the construction of IR systems by a series of programming assignments. Mathematical experience including basic probability is strongly desirable.

|  |
| --- |
| **Course Learning Outcomes (CLOs):** |
| At the end of the course students will be able to: |
| understand the common algorithms and techniques for information retrieval (document indexing and retrieval, query processing, etc. ) |
| use the quantitative evaluation methods for the IR systems and data mining techniques |
| implement a basic textual information retrieval system using Java or Python |
| Learn the popular probabilistic retrieval methods and ranking principles |
| apply information retrieval techniques to the problems of text clustering, recommendation systems, text classification etc. |

**Course Textbook**

[MRS] Introduction to Information Retrieval by Manning, Raghavan, and Schütze - available free online. (Online edition 2009)

**Additional references and books related to the course:**

[CMS] Search Engines by Croft, Metzler, and Strohman

**Tentative Weekly Schedule**

|  |  |  |
| --- | --- | --- |
| Week | Topics to be covered | Text Book Section |
| 1 | Key problems, Information need, Queries and documents, Matching scores | [MRS] Chapter 1 |
| 2 | **Inverted Index Construction**  Posting Lists, Dictionary  Distributed indexing, dynamic indexing, | [MRS] Chapter 4  (from 4.2 till 4.5) |
|  | **Zipf’s Law, Heap’s Law, Index Compression** | [MRS] Chapter 5  Section 5.1 and 5.3 |
| 3 | **Text Preprocessing**  Tokenization Stopping, stemming | [MRS] Chapter 2  Section 2.2 |
| 4 | **Retrieval Models (Vector Space Models)**  Vector-space model, Cosine Similarity, Tf-Idf, BM25 | [MRS] Chapter 6  Section 6.2 and 6.3 |
| 5 | **Retrieval Models ( Language Models)**  Smoothing Methods | [MRS] Chapter 12  12.1 to 12.3 |
| 6 |
| 7 | **Relevance Feedback** | [MRS] Chapter 9  Section 9.1.1, 9.1.3, 9.1.4, 9.1.6, 9.2 |
| 8 | **IR Evaluation/ Measures**  Ranking measures: R-prec, Mean Average Precision, nDCG, Reciprocal Rank | [MRS] Chapter 8  Section 8.1 to 8.4  (Interpolated precision is not included) |
| 9 | **Web Retrieval**  Link analysis, Markov Chains, PageRank | [MRS] Chapter 21  21.1, 21.2.1, 21.2.2 |
| 10 | **Clustering**  K-means clustering,  HAC, Evaluation of clustering | [MRS] Chapter 16  Section 16.3, 16.4 |
| 11 | **Clustering**  Hierarchal Agglomerative Clustering | [MRS] Chapter 17  Section 17.1 till 17.4 |
| 12 | **Text Classification**  Naive Bayes | [MRS] Chapter 13  Section 13.1 till 13.3 |
| 15 |  |  |
| 16 |  |  |

**(Tentative) Grading Criteria**

1. 2 Assignments (10%)
2. 3 Quizzes (10%)
3. Project (10%)
4. 2 Midterm Exam (25%)
5. Final Exam (45%)

**Course Policies**

1. Quizzes will be announced.
2. No makeup for missed quiz or assignment.
3. 80% attendance

**Plagiarism in Assignments**

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