Course Code: IT442

Introduction to Information retrieval

Instructor: Prasenjit Majumder

- **Introduction:**
  Goals and history of IR. The impact of the web on IR.

- **Basic IR Models:**
  Boolean and vector-space retrieval models; ranked retrieval; text-similarity metrics; TF-IDF (term frequency/inverse document frequency) weighting; cosine similarity.

- **Basic Tokenizing, Indexing, and Implementation of Vector-Space Retrieval:**
  Simple tokenizing, stop-word removal, and stemming; inverted indices;

- **Experimental Evaluation of IR:**
  Performance metrics: recall, precision, and F-measure; Evaluations on benchmark text collections.

- **Query Operations and Languages:**
  Relevance feedback; Query expansion; Query languages.

- **Text Representation:**
  Word statistics; Zipf's law; Porter stemmer; morphology; index term selection; using thesauri.

- **Web Search:**
  Search engines; spidering; metacrawlers; directed spidering; social network

- **Text Categorization and Clustering:**
  Categorization algorithms: Applications to information filtering and organization. Clustering algorithms: agglomerative clustering; k-means; Applications to web search and information organization.

- **Other retrieval Models** :

- **Text mining**

**Evaluation Scheme:**

Term paper: 40%
Class presentation 20%
Written Exam 40%