

Introduction to IR: History and motivation. Basic IR Models – Boolean and vector space models - Ranked Retrieval - Text Similarity Metrics. Basic Preprocessing: Tokenizing- stemming. Evaluations on benchmark text collections - Components of an information retrieval system. Indexing for IR: Inverted Indices - Postings lists –efficient processing with sparse vectors; Evaluation of IR - Performance metrics: recall, precision, and F-measure; Evaluations on benchmark text collections. Web Search: Search engines; spidering; metacrawlers; directed link analysis etc. Relevance in IR: Parametric or fielded search - Document zones - Vector space retrieval model - tf.idf weighting - queries as vectors - Computing scores in a complete search system - Efficient scoring and ranking – Evaluation in information retrieval : User happiness- Creating test collections: kappa measure- interjudge agreement - Relevance feedback and query expansion: Query expansion - Automatic thesaurus generation - Sense-based retrieval -. Document Classification and Clustering: Introduction to text classification -Latent Semantic Indexing. Recommender Systems - Collaborative filtering and content-based recommendation of documents and products. Introduction to Python NLTK and other toolkits.

TEXTBOOKS/REFERENCES:

1. C. Manning, P. Raghavan, and H. Schütze, “Introduction to Information Retrieval”, Cambridge University Press, 2008.
2. R. Baeza-Yates and B. Ribeiro Neto, “Modern Information Retrieval: The Concepts and Technology behind Search”, Second Edition, Addison Wesley, 2011.
3. David A. Grossman and Ophir Frieder “Information Retrieval: Algorithms and Heuristics”, Second Edition, Springer 2004.