

*Syllabus* : Introduction: concepts and terminology of information retrieval systems, Significance of information retrieval and storage, Information Retrieval Vs Information Extraction; Indexing: inverted files, encoding, Zipf's Law, compression, boolean queries; Fundamental IR models: Boolean, Vector Space, probabilistic, TFIDF, Okapi, language modeling, latent semantic indexing, query processing and refinement techniques; Performance Evaluation: precision, recall, F-measure; Classification: Rocchio, Naive Bayes, k-nearest neighbors, support vector machine; Clustering: partitioning methods, k-means clustering, hierarchical; Introduction to advanced topics: search, relevance feedback, ranking, query expansion.

*Texts* :

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schtze, Introduction to Information Retrieval, Cambridge University Press. 2008
2. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval, Addison Wesley, 1st edition, 1999.

*References* :

1. Soumen Chakrabarti, Mining the Web, Morgan-Kaufmann Publishers, 2002.
2. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, Corr. 2nd printing edition, 2009.
3. David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 2nd edition, 2004.
4. William B. Frakes, Ricardo Baeza-Yates, Information Retrieval Data Structures and Algorithms, Prentice Hall, 1992.
5. G. Salton, M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1986.
6. C. J. Van Rijsbergen, Information Retrieval, Butterworth-Heinemann; 2nd edition, 1979.