

**NIRMA UNIVERSITY**  
**Institute of Technology**  
**B. Tech. Computer Science and Engineering**  
**Semester – VI**  
**Department Elective-I**

L	T	P	C
2	0	2	3

<b>Course Code</b>	2CSDE53
<b>Course Title</b>	Information Retrieval Systems

**Course Outcomes:**

At the end of the course, students will be able to –

1. Correlate the concepts and various components of Information Retrieval (IR) systems
2. identify design and evaluation parameters for information retrieval systems
3. apply theoretical foundations for development of IR systems

**Syllabus:**

**Teaching  
Hours: 30**  
05

**Unit I**

**Introduction:** Concept and architecture of Information Retrieval systems, Boolean Retrieval, the term vocabulary and posting lists, text processing - tokenization, stop words removal, stemming, lemmatization, posting lists intersection via skip pointers, positional posting lists and phrase queries.

**Unit II**

**Data Structures:** Dictionaries and tolerant retrieval - wildcard queries, spelling correction, phonetic correction. Inverted indexing, index construction - types of indices, Index compression - Heap's law and Zipf's law, dictionary compression.

**Unit III**

**Scoring and Vector Space Models and Language Models:** Scoring, term weighting and vector space model, term frequency and weighting, TF-IDF weighting, Dot products, similarity measures for computing score, Evaluating IR systems, Relevance feedback - the Rocchio method, pseudo and indirect relevance feedback, query expansion, types of language models for information retrieval, the query likelihood model.

**Unit IV**

**Document Classification and Clustering:** Text classification using naive Bayesian method, Bernoulli and multinomial models, evaluation of text classification, other classification techniques

**Unit V**

**Web based IR:** Web crawling, web search and link analysis, Meta search engines, Multimedia IR - retrieving information from audio, video and images

29

**Self-Study:**

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

**Laboratory Work:**

Laboratory work will be based on the above syllabus with minimum 10 experiments to be incorporated.

**Suggested Readings<sup>^</sup>:**

1. Christopher D. Manning, PrabhakarRaghavan, HinrichSchutze, An Introduction to Information Retrieval, Cambridge University Press
2. D.A. Grossman, O. Frieder, Information Retrieval: Algorithms and Heuristics, Springer
3. W.B. Croft, J. Lafferty, Language Modeling for Information Retrieval, Springer
4. G. Kowalski, M.T. Maybury, Information Storage and Retrieval Systems, Springer
5. B. Croft, D. Metzler, T. Strohman, Information Retrieval in Practice, Pearson Education

L=Lecture, T=Tutorial, P=Practical, C=Credit

---

<sup>^</sup>this is not an exhaustive list