

Subject Name : Information Retrieval

Subject Code : IT 802-1 / CE 802-1

Teaching Scheme (Credits and Hours)

Teaching scheme				Total Credit	Evaluation Scheme					Total
L	T	P	Total		Theory		Mid Sem Exam	CIA	Pract.	
Hrs	Hrs	Hrs	Hrs		Hrs	Marks	Marks	Marks	Marks	
03	00	02	05	4	3	70	30	20	30	150

Learning Objectives:

- Learn to write code for text indexing and retrieval.
- Learn to evaluate information retrieval systems
- Learn to analyze textual and semi-structured data sets
- Learn to evaluate information retrieval systems
- Learn about text similarity measure
- Understanding about search engine
- Text Classification

Outline of the Course:

Sr. No	Title of the Unit	Minimum Hours
1	Overview of text retrieval systems	5
2	Retrieval models and implementation: Vector Space Models	6
3	Query expansion and feedback	5
4	Probabilistic models; statistical language models	8
5	Text classification & Text clustering	10
6	Web search basics, crawling, indexes, Link analysis	8
7	IR applications	3

Total hours (Theory): 45

Total hours (Lab): 30

Total hours: 75

Detailed Syllabus:

Sr. No	Topic	Lecture Hours	Weight age(%)
1	Overview of text retrieval systems <ul style="list-style-type: none">• Boolean retrieval• The term vocabulary and postings lists• Dictionaries and tolerant retrieval• Index construction and compression	5	12
2	Retrieval models and implementation: Vector Space Models <ul style="list-style-type: none">• Vector Space Model• TF-IDF Weight• Evaluation in information retrieval	6	15
3	Query expansion and feedback <ul style="list-style-type: none">• Relevance feedback• pseudo relevance feedback• Query Reformulation	5	12
4	Probabilistic models; statistical language models <ul style="list-style-type: none">• Okapi/BM25;• Language models• KL-divergence• Smoothing	8	15
5	Text classification & Text clustering <ul style="list-style-type: none">• The text classification problem• Naive Bayes text classification• k- nearest neighbors• Support vector Machine• Feature Selection• Vector-space clustering;• K-means algorithm• Hierarchical clustering• DBSCAN algorithm• PAM and PAMK• EM algorithm	10	22
6	Web search basics, crawling, indexes, Link analysis <ul style="list-style-type: none">• Web Characteristic• Crawling• Web As a graph• Page Rank• Hubs and Authorities	8	15

7	IR applications <ul style="list-style-type: none"> • Information extraction • Question answering • Opinion summarization • Social Network 	3	9
	Total	45	100

Instructional Method and Pedagogy:

- At the start of course, the course delivery pattern, prerequisite of the subject will be discussed.
- Lectures will be conducted with the aid of multi-media projector, black board, OHP etc.
- Attendance is compulsory in lecture and laboratory which carries 10 marks in overall evaluation.
- One internal exam will be conducted as a part of internal theory evaluation.
- Assignments based on the course content will be given to the students for each unit and will be evaluated at regular interval evaluation.
- Surprise tests/Quizzes/Seminar/tutorial will be conducted having a share of five marks in the overall internal evaluation.
- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Experiments shall be performed in the laboratory related to course contents.

Learning Outcome:

- To Understand Document as Vector
- Performance evolution metric for IR
- To understand search Engine functionality
- Various Supervised and Unsupervised learning Method

Text Book:

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>
- ChengXiang Zhai, Statistical Language Models for Information Retrieval (Synthesis Lectures Series on Human Language Technologies), Morgan & Claypool Publishers, 2008.
- <http://www.morganclaypool.com/doi/abs/10.2200/S00158ED1V01Y200811HLT001>

List of Project:

Sr. No	Name of Experiment
1	Implementation of various classification algorithm on text
2	Implementation of various Clustering algorithm on text
3	Implement Domain specific Search Engine
4	Social media analytic
5	Design and development of Question/Answering System
6	IR from Micro blog
7	<p>Various track at TREC 2015 conference (students will be encouraged to participate in such track)</p> <ul style="list-style-type: none">• Clinical Decision Support Track• Contextual Suggestion Track• Microblog Track• Temporal Summarization Track• Tasks Track
8	<p>Various track at CLEF 2015 Conference(students will be encouraged to participate in below track)</p> <ul style="list-style-type: none">• Linked Data Track• Tweet Contextualization track• Relevance Feedback Track