Kadi Sarva Vishwavidyalaya

Faculty of Engineering & Technology

Fourth Year Bachelor of Engineering (Computer/IT)

(To be Proposed For: Academic Year 2020-21)

Subject Code: CT803C-N	**Subject Title: Information Retrieval**
Pre-requisite	Linear Algebra, probability, data structure and algorithms

Teaching Scheme (Credits and Hours)

Teaching scheme				**Total Credit**	**Evaluation Scheme**					
L	**T**	**P**	**Total**		**Theory**		**Mid Sem Exam**	**CIA**	**Pract.**	**Total**
Hrs	**Hrs**	**Hrs**	**Hrs**		**Hrs**	**Marks**	**Marks**	**Marks**	**Marks**	**Marks**
04	**00**	**02**	**06**	**5**	**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 analyse textual and semi-structured data sets
- Learn to evaluate information retrieval systems
- Learn about text similarity measure
- Neural Information Retrieval
- Understanding about search engine

Outline of the Course:

Sr. No	**Title of the Unit**	**Minimum Hours**
1	Overview of text retrieval systems	6
2	Retrieval models and implementation: Vector Space Models	9
3	Probabilistic models; statistical language models	10
4	Query expansion and feedback	7
5	Neural Information Retrieval	20
6	Web search basics, crawling, indexes, Link analysis	7
7	IR applications	5

Total hours (Theory): 64
Total hours (Lab): 32
Total hours: 96

Kadi Sarva Vishwavidyalaya

Faculty of Engineering & Technology

Fourth Year Bachelor of Engineering (Computer/IT)

(To be Proposed For: Academic Year 2020-21)

Detailed Syllabus:

Sr. No	Topic	Lecture Hours	Weightage(%)
1	**Overview of text retrieval systems** • Boolean retrieval • The term vocabulary and postings lists • Dictionaries and tolerant retrieval • Index construction and compression	6	8
2	**Retrieval models and implementation: Vector Space Models** • Vector Space Model • TF-IDF Weight • Evaluation in information retrieval	9	15
3	**Probabilistic models; statistical language models** • Okapi/BM25; • Language models • KL-divergence • Smoothing	10	15
	Query expansion and feedback • Relevance feedback • pseudo relevance feedback • Query Reformulation	7	12
5	**Neural Information Retrieval** • Neural networks fundamentals • Learning to rank • Embeddings • Deep neural networks • Shallow neural methods for ranking • Deep neural methods for rankings • Deep neural methods for retrieval	20	33
6	**Web search basics, crawling, indexes, Link analysis** • Web Characteristic • Crawling • Web As a graph • Page Rank • Hubs and Authorities	7	12

7	**IR applications** • Information extraction • Question answering • Text summarization	5	5
	Total	**64**	**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:

- Text representation using vector
- Performance evolution metric for IR
- Understand search Engine functionality
- Get introduced with neural Information Retrieval

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
- Bhaskar Mitra and Nick Craswell, An Introduction to Neural Information Retrieval, Now publishers Inc.
- 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

Kadi Sarva Vishwavidyalaya

Faculty of Engineering & Technology

Fourth Year Bachelor of Engineering (Computer/IT)

(To be Proposed For: Academic Year 2020-21)

List of Experiments:

Sr. No	Name of Experiment
1	Text Representation by various weighting scheme
2	Introduction to Lucerne/terrier/Indri and Sample index creation in Java/Python.
3	Study various IR measure precision, Recall, F1-score, MAP, nDCG
4	Implement vector space model in R or Python
5	Implement Okapi BM25 model in R or python
6	Implement N-gram language model with different smoothing techniques
7	Implement LSA in R or Python
8	Implement Learning to Rank for text Retrieval
9	Implement deep neural network LSTM/CNN in R/Python
10	Implement rankings using deep neural methods
11	Various track at TREC 2020 conference (students will be encouraged to participate in such track) • Deep Learning Track • Incident Streams Track • Complex Answer Retrieval Track