Programme: M. Tech. (CSE)	Course Title: Information Retrieval and Web Search		Course Code:
Type of Course: Program Elective	Prerequisites: Mathematical Structures for Engineers (for PG) NOTE: Not open to UG students who have done IR course		Total Contact Hours: 40
Year/Semester: 1/Even	Lecture Hrs/Week: 3	Tutorial Hrs/Week: 0	Practical Hrs/Week: 0	Credits: 3

Learning Objective:

The growth of data is exponential in nature and the Internet is undisputedly the world's largest collection of information. In such a huge pool of web data, search engines are becoming the key technologies to extract useful material among the billions of available resources. This requires automated techniques to find, analyse, store, and extract relevant information from data. The course introduces standard concepts in information retrieval such as documents, queries, collections, and relevance. It also covers efficient indexing, retrieval techniques and their applications in recommender systems. Techniques used to retrieve useful information from repositories such as the Web are discussed.

Course outcomes (COs):

On completion of this course, the students will have the ability to:	Bloom's Level
CO-1	**Understand** the common algorithms and techniques for information retrieval	2
CO-2	**Understand** quantitative evaluation methods and **apply** retrieval methods and ranking principles for **analysis** of retrieved data.	2, 3, 4
CO-3	**Apply** text classification and clustering algorithms to group documents	3
CO-4	**Understand** and **Analyse** the techniques and algorithms existing in practical retrieval in web search engines and recommender systems.	2, 4

Course Topics	Lecture Hours	
UNIT – I Basic IR Models, Tokenizing, Indexing, and Implementation of Vector-Space Retrieval	9	9
1.1 Boolean and vector-space retrieval models; Ranked retrieval; Text similarity metrics.	3	

1.2 TF-IDF (term frequency/inverse document frequency) weighting; Cosine similarity.	**2**	
1.3 Simple tokenizing, stop-word removal, and stemming; inverted indices; efficient processing with sparse vectors, index construction and compression	**4**	
UNIT – II Experimental Evaluation of IR, Query Operations and Languages	**9**	
2.1 Performance metrics: recall, precision, and F- measure	**2**	**9**
2.2 Evaluations on benchmark text collections	**3**	
2.3 Relevance feedback; Query expansion; Query languages	**4**	
UNIT – III: Text Representation, Categorization and Clustering	**9**	
3.1 Word statistics; Zipf's law; Porter stemmer; Morphology; Index term selection; Using thesauri. Metadata and markup languages (SGML, HTML, XML)	3	**9**
3.2 Categorization using Naive Bayes, Decision trees	3	
3.3 Clustering using Agglomerative clustering; k-means; Expectation maximization (EM); Latent Semantic Analysis	3	
UNIT – IV: Web Search	**9**	
4.1 Search engines; Spidering; Metacrawlers	**3**	**9**
4.2 Link analysis; Shopping agents	**3**	
4.3 Recent concepts and algorithms	**3**	
UNIT-V: Recommender Systems	**4**	**4**
5.1 Collaborative and Content based recommendation of documents and products	**4**	

Textbook references:

Text Books:

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, *Introduction to Information Retrieval,* Cambridge University Press, 2008.
2. Soumen Chakrabarti, *Mining the web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann, 2002.

Reference Books:

1. D. Grossman and O. Frieder, *Information Retrieval: Algorithms and Heuristics*, 2nd Edition, Springer, 2004
2. R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, ACM Press, 1999
3. W. B. Croft, D. Metzler, and T. Strohman, *Search Engines – Information Retrieval in Practice*, Cambridge University Press, 2009
4. S. Büttcher, C. Clarke, and G. Cormack, *Information Retrieval: Implementing and Evaluating Search Engines*, MIT Press, 2010
5. Charu C. Aggarwal*, Recommnder Systems,* Springer, 2016

Other resources:

1. Hugh E. Williams, Justin Zobel, Dirk bahle, "Fast phrase querying with combined indexes", ACM Trans. Inf. Syst., vol. 22, no. 4, pp. 573-594, 2004.
2. Hao Yan, Shuai Ding, Torsten Suel, "Inverted Index Compression and Query Processing with Optimized Document Ordering", Proceedings of the 18th International Conference on World Wide Web, pp. 401-410, 2009.
3. How to write a spelling corrector, http://norvig.com/spell-correct.html
4. Paolo Boldi, Sebastian Vigna, "The Webgraph Framework I: Compression Techniques, "Proceedings of the Thirteenth International World Wide Web Conference" pp. 595-601, 2003

Evaluation Method	
Item	**Weightage (%)**
Quiz / Assignments	30
Project (Part A+ Part B)	20
Midterm	20
Final Examination	30

*Please note, as per the existing institute's attendance policy the student should have a minimum of 75% attendance. Students who fail to attend a minimum of 75% lectures will be debarred from the End Term/Final/Comprehensive examination.

CO and PO Correlation Matrix

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	1	2													
CO2	2	2		2	3								2		2
CO3	3	2	1	2	2								2		
CO4		3	1	3		1					1	2	3		3

Last Updated On: 04ᵗʰ October 2021

Prepared and Updated By: Preety Singh and Suvidha Tripathi

Approved By: