

Savitribai Phule Pune University
Fourth Year of Artificial Intelligence and Data Science (2020 Course)
Elective IV 417524(B): Information Retrieval

Teaching Scheme: TH: 03 Hours/Week	Credit 03	Examination Scheme: In-Sem (Paper): 30 Marks End-Sem (Paper): 70 Marks
---	----------------------------	---

Prerequisites Courses: Data Structures and Algorithms (210252), Database management systems(310241)

Companion Course: Computer Laboratory II (417526)

Course Objectives:

- To understand the basics of Information Retrieval
- To understand the concepts of Indexing & Query Processing for Information Retrieval
- To provide comprehensive details about various Evaluation methods
- To understand the different methods of Text Classification and Clustering
- To understand various search engine system operations and web structures
- To understand various applications of Information Retrieval

Course Outcomes:

After completion of the course, learners should be able to-

- CO1:** Understand the concept of Information Retrieval
CO2: Use an indexing approach for retrieval of documents
CO3: Evaluate and analyze the retrieved information
CO4: Apply appropriate method of Text Classification and Clustering
CO5: Design and implement innovative features in search engines
CO6: Analyze different real-life application of Information Retrieval

Course Contents

Unit I	Introduction	06 Hours
Introduction to information retrieval, Major challenges in IR, Features of an IR system, components of an IR model, IR system block diagram, Boolean retrieval, Information versus Data Retrieval, Text categorization, IR processes and fields, Vector Model, Probabilistic Model and Latent Semantic Indexing Model.		
#Exemplar/Case Studies	Visualization and Information Retrieval Techniques for Network Intrusion Detection	
*Mapping of Course Outcomes for Unit I	CO1	
Unit II	Dictionaries and Query Processing	06 Hours
Components of Index, Index Life Cycle, Static Inverted Index, Dictionaries-Types (Sort Based, Hash Based, Interleaving & Posting Lists), Index Construction (In memory, Sort Based, Merge Based, Disk Based Index Construction), Dynamic Indexing, Query Processing for Ranked Retrieval, Document at a Time Query Processing, Term at a Time Query Processing, Pre-computing Score Contributions, Impact Ordering, Query Optimization.		
#Exemplar/Case Studies	Matching of the searched statement with the database which is already stored	

*Mapping of Course Outcomes for Unit II		CO2
Unit III	Probabilistic Retrieval and Language Modelling related methods	06 Hours
<p>Probabilistic Retrieval: Review of Basic Probability Theory; The Probability Ranking Principle: The 1/0 loss case, the PRP with retrieval costs; The Binary Independence Model; Term Frequency; An appraisal and some extensions: An appraisal of probabilistic models, tree-structured dependencies between terms, Okapi BM25: a non-binary model, Bayesian network approaches to IR, Relevance Feedback, Field Weights:BM25F.</p> <p>Language models for information retrieval: generating queries from documents; Language models: finite automata and language models; types of language models; multinomial distributions over words; Ranking with Language Models; Divergence from Randomness, Passage Retrieval, and Ranking.</p>		
#Exemplar/Case Studies		A Comparative Study of Probabilistic and Language Models for Information Retrieval
*Mapping of Course Outcomes for Unit III		CO3
Unit IV	Text classification & Text clustering	06 Hours
<p>Text Classification: Introduction to Text Classification, Naïves Bayes Model, K Nearest neighbor, spam filtering, Support Vector Machine Classifier, Vector Space classification using hyperplanes, kernel function.</p> <p>Text Clustering: Clustering vs Classification, partitioning methods. Clustering Algorithms: k-means clustering, Agglomerative hierarchical clustering, Expectation Maximization, Mixture of Gaussians Model</p>		
#Exemplar/Case Studies		Case Study: Improving Document Organization and Retrieval in a Digital Library
*Mapping of Course Outcomes for Unit IV		CO4
Unit V	Web Retrieval and Web Crawling	06 Hours
<p>Parallel information retrieval: Parallel query processing, Mapreduce</p> <p>Web Retrieval: Search Engine Architectures, Cluster based Architecture, Distributed Architectures, Search Engine Ranking. Link based Ranking, Page Ranking Algorithm, Simple Ranking Functions and Evaluations.</p> <p>Web Crawler: Web Crawler structure, Web crawler libraries, Python Scrapy, BeautifulSoup, Applications</p>		
#Exemplar/Case Studies		Study of Google Map/ Geogusser information retrieval
*Mapping of Course Outcomes for Unit V		CO5
Unit VI	IR applications	06 Hours
<p>Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval</p> <p>Recommender System: Collaborative Filtering, Content Based Recommendation, Knowledge Based Recommendation Information Extraction and Integration: Extracting Data from Text. Semantic Web, Collecting and Integrating Specialized Information on the web.</p>		

#Exemplar/Case Studies	Demonstrate Collaborative filtering using any datasets to recommend items to users
*Mapping of Course Outcomes for Unit VI	CO6

Learning Resources

Text Books:

1. C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, 2008, ISBN: 9780521865715
2. S. Buttcher, C. Clarke, and G. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines" MIT Press, 2010, ISBN: 0-408-70929-4
3. Bruce Croft, Donald Metzler and Trevor Strohman, "Search Engines: Information Retrieval in Practice", 1st edition, Addison Wesley, 2009, ISBN: 9780135756324
4. Jannach D., Zanker M. and FelFering A., "Recommender Systems: An Introduction", 1st edition, Cambridge University Press

Reference Books:

1. Manouselis N., Drachsler H., Verbert K., Duval E., "Recommender Systems For Learning", 1st edition, Springer, 2013
2. G. Kowalski, M.T. Maybury, "Information storage and Retrieval System", Springer, 2005
3. W.B. Croft, J. Lafferty, "Language Modeling for Information Retrieval", Springer, 2003

e-Resources:

1. Information Retrieval- <http://www.informationretrieval.org>

MOOC Courses:

1. <https://www.youtube.com/watch?v=fFxpSmyICwI>
2. <https://www.youtube.com/watch?v=X5GvBh4qY0s>

The CO-PO Mapping Matrix

CO/ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	1	–	–	–	–	–	–	–	–	–	–	–
CO2	–	1	2	–	–	–	–	–	–	–	–	–
CO3	1	2	1	1	1	–	–	–	–	–	–	2
CO4	1	–	1	1	–	–	–	–	1	–	–	1
CO5	1	–	–	–	–	–	–	1	1	–	–	1
CO6	–	–	–	–	1	–	–	–	–	–	–	1