



SRI VENKATESWARA COLLEGE OF ENGINEERING

COURSE DELIVERY PLAN - THEORY

Page 1 of 6

Department of Information Technology		LP: Sub Code
B.E/B.Tech/M.E/M.Tech : <u>Information Technology</u> Regulation: 2018		Rev. No:
PG Specialisation : _____		Date:
Sub. Code / Sub. Name : IT18011- Information Retrieval		
Unit : 1		

Unit Syllabus: UNIT 1 INTRODUCTION IR Model – Boolean Retrieval –The term Vocabulary and posting lists -Tokenization – Stop words – Stemming – Inverted index – Skip pointers – Phrase Queries- Dictionaries and tolerant retrieval.

Objective: Build an open source search engine framework and explore its capabilities

Session No *	Topics to be covered	Ref	Teaching Aids
1	Introduction to IR	RB-1	PPT/Online
2	IR Model- Boolean Retrieval	TB-1,Page 1-17	PPT/Online
3	IR Model - Boolean Retrieval	TB-1,Page 1-17	PPT/Online
4	The term Vocabulary and Posting lists	TB-1,Page 19-20	PPT/Online
5	Tokenization and Stop words	TB-1,Page 22	PPT/Online
6	Stemming	TB-1,Page 27	PPT/Online
7	Inverted index – Skip Pointers	TB-1,Page 36	PPT/Online
8	Phrase Queries	TB-1,Page 39	PPT/Online
9	Dictionaries and tolerant retrieval.	TB-1,Page 49	PPT/Online
Content beyond syllabus covered (if any):			

* Session duration: 50 minutes



Sub. Code / Sub. Name: IT18011- Information Retrieval
Unit : II : Language Models and Indexing

Unit Syllabus : Language Models for IR – Index Construction – Index Compression – Scoring – Term weighing – vector space model – Computing scores in a complete search system

Objective: Classify documents in different ways and discuss its effect on similarity calculations and on search

Session No *	Topics to be covered	Ref	Teaching Aids
10	Language Models for IR	TB-1, Page 237	PPT/Online
11	Language Models for IR	TB-1, Page 237	PPT/Online
12	Index Construction	TB-1, Page 67	PPT/Online
13	Index Compression	TB-1, Page 85	PPT/Online
14	Scoring	TB-1, Page 109	PPT/Online
15	Term weighing	TB-1, Page 109	PPT/Online
16	vector space model	TB-1, Page 109	PPT/Online
17	Computing scores in a complete search system	TB-1, Page 135	PPT/Online
18	Computing scores in a complete search system	TB-1, Page 135	PPT/Online

Content beyond syllabus covered (if any):

* Session duration: 50 mins



Sub. Code / Sub. Name: TT18011- Information Retrieval
Unit : III : EVALUATION

Unit Syllabus : Evaluation in Information Retrieval - Relevance Feedback and Query Expansion - XML Retrieval - Probabilistic Information Retrieval
Objective: Compare practical algorithms of textual document indexing, relevant ranking, web mining, text analytics and their performance evaluations.

Session No *	Topics to be covered	Ref	Teaching Aids
19	Evaluation in Information Retrieval	TB-1,Page 151	PPT/Online
20	Evaluation in Information Retrieval	TB-1,Page 151	PPT/Online
21	Relevance Feedback and Query Expansion	TB-1,Page 177	PPT/Online
22	Relevance Feedback and Query Expansion	TB-1,Page 177	PPT/Online
23	Relevance Feedback and Query Expansion	TB-1,Page 177	PPT/Online
24	XML Retrieval	TB-1,Page 195	PPT/Online
25	XML Retrieval	TB-1,Page 195	PPT/Online
26	Probabilistic Information Retrieval	TB-1,Page 219	PPT/Online
27	Probabilistic Information Retrieval	TB-1,Page 219	PPT/Online

Content beyond syllabus covered (if any):

* Session duration: 50 mins



Sub. Code / Sub. Name: IT18011- Information Retrieval

Unit : IV – Crawling and Link Analysis.

Unit Syllabus : Web Search basics – Search Advertising – Duplicate Detection – Web Crawling and Indices - Crawling architecture – Distributed Crawling - Link Analysis – Web as graph – Page Rank
 Objective: Develop the necessary experience to design, and implement applications using Information Retrieval systems

Session No *	Topics to be covered	Ref	Teaching Aids
28	Web Search basics	TB-1,Page 421	PPT/Online
29	Search Advertising	TB-1,Page 431	PPT/Online
30	Duplicate Detection	TB-1,Page 431	PPT/Online
31	Web Crawling and Indices	TB-1,Page 443	PPT/Online
32	Crawling architecture	TB-1,Page 443	PPT/Online
33	Distributed Crawling	TB-1,Page 485	PPT/Online
34	Link Analysis	TB-1,Page 461	PPT/Online
35	Web as graph	TB-1,Page 462	PPT/Online
36	Page Rank	TB-1,Page 464	PPT/Online

Content beyond syllabus covered (if any):

* Session duration: 50 mins



Sub. Code / Sub. Name: IT18011- Information Retrieval

Unit : V – Applications.

Unit Syllabus : Information Extraction – Automatic Text Summarization - Question Answering Systems - Spam Filtering – Document Classification – Cross Lingual retrieval – Compound Term Processing
Objective: Design and implement an innovative feature in a search engine

Session No *	Topics to be covered	Ref	Teaching Aids
37	Information Extraction	TB-2, Chapter 11.2	PPT/Online
38	Automatic Text Summarization	TB-2, Page 368	PPT/Online
39	Automatic Text Summarization	TB-2, Page 368	PPT/Online
40	Question Answering Systems	TB-2, Page 359	PPT/Online
41	Spam Filtering	TB-1, Page 334	PPT/Online
42	Document Classification	TB-2, Page 332	PPT/Online
43	Cross Lingual retrieval	TB-2, Chapter 10.6	PPT/Online
44	Compound Term Processing	Online. https://en.wikipedia.org/wiki/Compound	PPT/Online
45	Compound Term Processing	https://en.wikipedia.org/wiki/Compound-term_processing	PPT/Online
Content beyond syllabus covered (if any):			

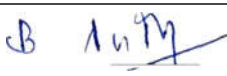

* Session duration: 50 mins

**TEXT BOOKS:**

1. C. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
2. "Natural Language Processing And Information Retrieval", Tanveer Siddiqui and U. S. Tiwary, Oxford University Press.

REFERENCES:

1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, —Modern Information Retrieval
2. Concepts and Technology behind Search, Second Edition, ACM Press Books, 2011.
3. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, —Information Retrieval:
4. Implementing and Evaluating Search Engines, The MIT Press, 2010.

	Prepared by	Approved by
Signature		
Name	Dr. B. Prakash	Dr. V. Vidhya
Designation	Asst.Prof	Prof & Head
Date	24.07.2021	24.07.2021
Remarks *:		

* If the same lesson plan is followed in the subsequent semester/year it should be mentioned and signed by the Faculty and the HOD