THIRD-YEAR OF BACHELOR OF COMPUTER SCIENCE REVISED SYLLABUS ACCORDING TO CBCS

COURSE TITLE: INFORMATION RETRIEVAL

SEMESTER-V, W.E.F. 2021-2022

Recommended by the Board of Studies in Computer Science
And
Approved by the Academic Council
Devrukh Shikshan Prasarak Mandal's
Nya. Tatyasaheb Athalye Arts, Ved. S. R. Sapre Commerce, and
Vid. Dadasaheb Pitre Science College (Autonomous), Devrukh.
Tal. Sangameshwar, Dist. Ratnagiri-415804, Maharashtra,
India

Name of the Implementing Institute	:	Nya. Tatyasaheb Athalye Arts, Ved. S. R. Sapre Commerce, and Vid. Dadasaheb Pitre Science College (Autonomous), Devrukh. Tal. Sangameshwar, Dist. Ratnagiri-415804,
Name of the Parent University	:	University of Mumbai
Name of the Programme	:	Bachelor of Science
Name of the Department	:	Computer Science
Name of the Class	:	Third Year
Semester	:	Six
No. of Credits	:	03
Title of the Course	:	Information Retrieval
Course Code	:	USCST64
Name of the Vertical	:	Elective II
Eligibility for Admission	:	Any 12th Pass seeking Admission to Degree Programme in adherence to Rules and Regulations of the University of Mumbai and Government of Maharashtra
Passing Marks	:	40%
Mode of Assessment	:	Formative and Summative
Level	:	UG
Pattern of Marks Distribution for TE and CIA	:	70:30
Status	:	CBCS
To be implemented from Academic Year	:	2021-2022
Ordinances /Regulations (if any)		

Syllabus for Third Year of Bachelor of Science in Computer Science

(With effect from the academic year 2021-2022)

SEMESTER-VI   **Paper No.– 4**

Course Title: Information Retrieval   **No. of Credits - 03**

Type of Vertical: Elective II   **COURSE CODE: USCST64**

Learning Outcomes Based on BLOOM's Taxonomy:

After completing the course, the learner will be able to…		
Course Learning Outcome No.	Blooms Taxonomy	Course Learning Outcome
CO-01	Understand	After completion of this course, learner should get an understanding of the field of information retrieval and its relationship to search engines.
CO-02	Understand	It will give the learner an understanding to apply information retrieval models.

Syllabus for Third Year of Bachelor of Science in Computer Science

(With effect from the academic year 2021-2022)

SEMESTER-VI   **Paper No.–4**

Course Title: Information Retrieval   **No. of Credits - 03**

Type of Vertical: Elective II   **COURSE CODE: USCST64**

	COURSE CONTENT		
Unit No.	**Content**	**Credits**	**No. of Lectures**
I	Introduction to Information Retrieval: Introduction, History of IR, Components of IR, and Issues related to IR, Boolean retrieval, Dictionaries and tolerant retrieval.	**01**	**15**
II	Link Analysis and Specialized Search: Link Analysis, hubs and authorities, Page Rank and HITS algorithms, Similarity, Hadoop & Map Reduce, Evaluation, Personalized search, Collaborative filtering and content-based recommendation of documents and products, handling ─invisible‖ Web, Snippet generation, Summarization, Question Answering, CrossLingual Retrieval.	**01**	**15**
III	Web Search Engine: Web search overview, web structure, the user, paid placement, search engine optimization/spam, Web size measurement, search engine optimization/spam, Web Search Architectures. XML retrieval: Basic XML concepts, Challenges in XML retrieval, A vector space model for XML retrieval, Evaluation of XML retrieval, Text-centric versus data-centric XML retrieval.	**01**	**15**
	Total	03	45

Required Previous Knowledge

Students should know basic concepts related to computer and computer handling

Access to the Course

The course is available for all the students admitted for Bachelor of Science (Computer Science).

Forms of Assessment

The assessment of the course will be of Diagnostic, Formative and Summative type. At the beginning of the course diagnostic assessment will be carried out. The formative assessment will be used for the Continuous Internal Evaluation whereas the summative assessment will be conducted at the end of the term. The weightage for formative and summative assessment will be 60:40. The detailed pattern is as given below.

Semester End Evaluation (70 Marks)
Question Paper Pattern
Time: 2:30 hours

Question No.	Unit/s	Question Pattern	Marks
Q.1	I ,II &III	MCQ/Fill in the blanks/One line sentence	10
Q.2	I	Descriptive Questions	20
Q.3	II	Descriptive Questions	20
Q4.	III	Descriptive Questions	20
		Total	**70**

Internal evaluation (30 Marks)

Sr. No.	Description	Marks
1	Classroom Tests	10
2	Project/ Viva/ Presentations/ Assignments	10
3	Attendance	10
	Total	**30**

Grading Scale

10 points grading scale will be used. The grading scale used is O to F. Grade O is the highest passing grade on the grading scale, and grade F is a fail. The Board of Examinations of the college reserves the right to change the grading scale.

Reference book:

- Introduction to Information Retrieval, C. Manning, P. Raghavan, and H. Schütze, Cambridge University Press, 2008
- Modern Information Retrieval: The Concepts and Technology behind Search, Ricardo Baeza -Yates and Berthier Ribeiro – Neto, 2nd Edition, ACM Press Books 2011.
- Search Engines: Information Retrieval in Practice, Bruce Croft, Donald Metzler and Trevor Strohman, 1st Edition, Pearson, 2009.

Text book:
- Techmax publication book

Additional References:
- Information Retrieval Implementing and Evaluating Search Engines, Stefan Büttcher, Charles L. A. Clarke and Gordon V. Cormack, The MIT Press; Reprint edition (February 12, 2016),2nd

Course: USCSP69	Practical of USCST64 (Credits : 1, Lectures/Week: 3)
USCSP69	Practical may be done using software/tools like Python / Java / Hadoop 1. Write a program to demonstrate bitwise operation. 2. Implement Page Rank Algorithm. 3. Implement Dynamic programming algorithm for computing the edit distance between strings s1 and s2. (Hint. Levenshtein Distance) 4. Write a program to Compute Similarity between two text documents. 5. Write a map-reduce program to count the number of occurrences of each alphabetic character in the given dataset. The count for each letter should be case-insensitive (i.e., include both upper-case and lower-case versions of the letter; Ignore non-alphabetic characters). 6. Implement a basic IR system using Lucene. 7. Write a program for Pre-processing of a Text Document: stop word removal. 8. Write a program for mining Twitter to identify tweets for a specific period and identify trends and named entities. 9. Write a program to implement simple web crawler. 10. Write a program to parse XML text, generate Web graph and compute topic specific page rank