

COURSE TITLE: **INFORMATION RETRIEVAL**

A Program Elective Course for CSE

Faculty: **Dr. Rajendra Prasath**, Associate Professor
Faculty of Computer Science and Engineering
Indian Institute of Information Technology Sri City
Andhra Pradesh - 517646, India.

Email: rajendra.prasath@iiits.in

Phone: +91 76020 80978 (personal)

Course WEBSITE: <https://www.2power3.com/rajendra/>

Then go to Menu -> Courses -> Monsoon 2020: Information Retrieval (Program Elective Course) and explore Menu (Horizontal) Options: Contents, Text Books, Slides, Tutorials, Assignments, Exams, Practice

1. COURSE OUTLINE:

Information retrieval covers the tasks of indexing, searching, and recalling data, particularly text or other unstructured forms. It has an important role to play in a large number of applications viz., digital libraries, office automation, Internet, and e-commerce.

Information Retrieval (IR) is a program elective course offered to the students of CSE at IIIT Sri City. IR handles various search problems in any area of computing.

2. COUSE OBJECTIVES:

This introductory course on Information Retrieval will illustrate information processing approaches that could support the extraction, storage, and retrieval of information. These concepts could be adapted with suitable data structures for various problems that need scaling of information in various areas of computing.

The aim of the course is to study theoretical aspects as well as implementation issues of classical and modern retrieval problems.

This course is offered to those who are familiar with at least one programming language and the usage of data structures in that programming language. This course has a rigorous practice-oriented application development component.

3. BOOKS:

TEXTBOOKS:

1) Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley Longman Publishing, Boston, MA, USA, 1999.

2) Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, An Introduction to Information Retrieval, Cambridge University Press, Cambridge, England, 2009

3) William B. Frakes and Ricardo Baeza-Yates (Eds.), Information Retrieval: Data Structures and Algorithms, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1992.

4) State-of-the-art research papers from SIGIR, WWW, KDD, ECIR and AIRS conferences

4. COURSE OUTLINE:

This course is divided into the following modules:

M1: Fundamentals - Basics of Information Retrieval

- Introduction
- Boolean retrieval
- The term vocabulary & postings lists
- Dictionaries and tolerant retrieval
- Index construction

The above topics would be covered before the Mid Semester - 1 examination

M2: Modern Information Retrieval

- Index compression
- Scoring, term weighting & the vector space model
- Computing scores in a complete search system
- Evaluation in information retrieval

M3: Approaches in Information Retrieval

- Relevance feedback & query expansion
- XML retrieval
- Probabilistic information retrieval
- Language models for information retrieval
- Information Extraction

The above topics would be covered before the Mid Semester - 2 examinations

M4: Advances in Information Retrieval

- Text classification & Naive Bayes
- Vector space classification

Flat clustering
Hierarchical clustering

M5: Web Search and Recommendations

Recommender Systems
Web search basics
Web crawling and indexes
Link analysis

The above topics would be covered before the End Semester examinations

M6: Applications of Information Retrieval

Scalable Applications of IR
Graphs – Massive Web graph / Scale free Graphs
Path estimations between given two locations
Scalable Graph Examples: Small World Networks
Code Search
Meta Search Engines – Combining Search results of 5 search engines
Handling of data from Forums and Blogs
Argumentation Mining
Mining Unstructured Text Data
News Document Retrieval
Scientific Documents Retrieval
Smart Data Analytics from Unstructured Text Data
Understanding Text in Health domain

5. PRE-REQUISITES:

Any course on the Design and Analysis of Algorithms and Computer Programming, or its equivalent:

- a) Must have sufficient exposure to various algorithm analysis techniques in computing
- b) Must be aware of text processing and search engines
- c) Must have a good knowledge in basic probability theory, and ability to handle different data structures

6. COURSE GRADING:

Course grades will be based on the following weightage pattern: This is purely indicative and necessary modification will be made by the faculty depending on the intensity of the different work assigned for various tasks.

- a) EXAMINATIONS: 50%

Mid Semester - 1 Exam : 10%
Mid Semester - 2 Exam : 15%
End Semester Exam : 25%

b) TAKE HOME ASSIGNMENTS: 20%

- i) Students are given a set of problems every week.
- ii) The assignment must be solved by individuals
- iii) All assignments must be finished before the deadline specified for that set of problems
- iv) All Assignments are COMPULSORY

NOTE:

- i) If you fail to submit your solution, you will get "0"
- ii) Solutions would be cross-checked!!
- iii) Solutions submitted after the deadline will not be considered for evaluation.
- iv) Submission Procedure would be given to the students.

c) OPEN SECTION: 30%

This weightage is to provide different options for exploring new and innovative solutions for a given problem using computational geometry techniques.

The total weightage would be distributed in the following way:

Total Weightage (100%) = Take Home Assignments (20%) + Exams (50%) + Best Solutions (10%) + Specific Task Completion (20%)

Remember that we should follow the Academic Code of Conduct and students must explore severe PENALTIES.

7. PENALTIES:

Every Student is expected to strictly follow a fair Academic Code of Conduct to avoid severe penalties. Penalties would be heavy for those who involve in:

- Copy and Pasting of the Assignment
- Plagiarism (copied from your neighbor or friend – in this case, both will get "0" marks for that specific take-home assignments)
- If the candidate is unable to explain his own approach, it would be considered as a "copied case" !!
- Any other unfair means of completing the assignments

8. COLLABORATIVE LEARNING:

We plan to identify three sections of students:

- Perspective Students (having CGPA above 8.5 and above)
- Promising Students (having CGPA above 6.5 and less than 8.5)
- Needy Students (having CGPA less than 6.5)

The first two groups of students will enable and grow a culture of collaborative learning by helping the needy students. Such involvement would be rewarded.

9. ASSISTANCE:

- a) Students can contact me during the working hours with a prior appointment. In turn, students may also post their questions to me at any time over email.
- b) Students requiring more time or lengthy discussions may meet me in person during the working days by finding a suitable time slot or with a prior appointment.
- c) Course TAs would assist you to clear your doubts. Also you could get assistance from Perspective or Promising students to clear your doubts.
- d) The best way to reach me: Just send an email to my institute email id.