

|                              |                                        |
|------------------------------|----------------------------------------|
| <b>Branch Name:</b>          | Artificial Intelligence & Data Science |
| <b>Program Code:</b>         | ET108                                  |
| <b>Semester/Year:</b>        | Semester VII/ Fourth Year              |
| <b>Course Name:</b>          | Information Retrieval                  |
| <b>Course Code:</b>          | 1ET1080702                             |
| <b>Pre-requisite Course:</b> | Basics of Machine Learning             |

**Course Objective:**

- Learn to write code for text indexing and retrieval.
- Learn to evaluate information retrieval systems
- Learn to analyze textual and semi-structured data sets
- Learn to evaluate information retrieval systems
- Learn about text similarity measure
- Understanding about search engine
- Text Classification

**Teaching and Examination Scheme:**

| Teaching Scheme Hours per week) |                 |                  |        | Evaluation Scheme (Marks) |                          |                          |                          |                  |
|---------------------------------|-----------------|------------------|--------|---------------------------|--------------------------|--------------------------|--------------------------|------------------|
| Lecture<br>(L)                  | Tutorial<br>(T) | Practical<br>(P) | Credit | Theory (Marks)            |                          | Practical (Marks)        |                          | Total<br>(Marks) |
|                                 |                 |                  |        | University<br>Assessment  | Continuous<br>Assessment | University<br>Assessment | Continuous<br>Assessment |                  |
| 04                              | -               | 02               | 05     | 70                        | 30                       | 30                       | 20                       | 150              |

**Subject Contents:**

| Sr.<br>No. | Topic                                                                                                                                                                                                                                                                                                   | Total<br>Hrs. | Weightage<br>% |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------|
| 1          | <b>Overview of text retrieval systems</b><br>Boolean retrieval, The term vocabulary and postings lists, Dictionaries and tolerant retrieval, Index construction and compression                                                                                                                         | 5             | 12             |
| 2          | <b>Retrieval models and implementation: Vector Space Models</b><br>Vector Space Model, TF-IDF Weight, Evaluation in information retrieval                                                                                                                                                               | 6             | 15             |
| 3          | <b>Query expansion and feedback</b><br>Relevance feedback, pseudo relevance feedback, Query Reformulation                                                                                                                                                                                               | 5             | 12             |
| 4          | <b>Probabilistic models; statistical language models</b><br>Okapi/BM25; Language model, KL-divergence, Smoothing                                                                                                                                                                                        | 8             | 15             |
| 5          | <b>Text classification &amp; Text clustering</b><br>The text classification problem, Naive Bayes text classification, k- nearest neighbors, Support vector Machine, Feature Selection, Vector-space clustering; K-means algorithm, Hierarchical clustering, DBSCAN algorithm PAM and PAMK, EM algorithm | 10            | 22             |
| 6          | <b>Web search basics, crawling, indexes, Link analysis</b><br>Web Characteristic, Crawling, Web As a graph, Page Rank, Hubs and Authorities                                                                                                                                                             | 8             | 15             |
| 7          | <b>IR applications</b><br>Information extraction, Question answering, Opinion summarization, Social Network                                                                                                                                                                                             | 3             | 9              |

**List of Reference Books:**

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>
2. ChengXiang Zhai, Statistical Language Models for Information Retrieval (Synthesis Lectures Series on Human Language Technologies), Morgan & Claypool Publishers, 2008.
3. <http://www.morganclaypool.com/doi/abs/10.2200/S00158ED1V01Y200811HLT001>

**List of Suggested titles of Experiments (If Any):**

1. Implementation of various classification algorithm on text
2. Implementation of various Clustering algorithm on text
3. Implement Domain specific Search Engine
4. Social media analytic
5. Design and development of Question/Answering System
6. IR from Micro blog
7. Various track at TREC 2015 conference (students will be encouraged to participate in such track)
  - Clinical Decision Support Track
  - Contextual Suggestion Track
  - Microblog Track
  - Temporal Summarization Track
  - Tasks Track
8. Various track at CLEF 2015 Conference (students will be encouraged to participate in below track)
  - Linked Data Track
  - Tweet Contextualization track
  - Relevance Feedback Track

**Course Learning Outcomes (CLO): On completion of this course, the students will be able to:**

| CLO  | Description                                                    | Bloom's Taxonomy Level             |
|------|----------------------------------------------------------------|------------------------------------|
| CLO1 | Understand comprehend the key concepts of clustering algorithm | 1. Remembering<br>2. Understanding |
| CLO2 | Learn representation of Various track at CLEF 2015             | 2 Understanding<br>5. Evaluate     |
| CLO3 | Learn IR from Micro blog                                       | 4. Analyze                         |
| CLO4 | Learn Implement Domain specific Search Engine                  | 2 Understanding<br>5. Evaluate     |

**Mapping of CLOs with POs & PSOs**

| Course Learning Outcomes | Program Outcomes (POs) |     |     |     |     |     |     |     |     |      |      |      | Program Specific Outcomes(PSOs) |      |      |      |
|--------------------------|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|---------------------------------|------|------|------|
|                          | PO1                    | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1                            | PSO2 | PSO3 | PSO4 |
| CLO1                     |                        |     |     |     | L   | H   |     |     |     |      |      | H    | H                               | M    | L    |      |
| CLO2                     | H                      |     | M   | M   | L   |     |     |     |     | L    |      | M    |                                 | M    | L    |      |
| CLO3                     |                        | M   | L   |     | L   |     |     |     |     |      |      |      |                                 | H    |      | L    |
| CLO4                     | H                      | M   |     |     | L   |     |     | L   |     |      |      |      | L                               |      | M    | H    |