# Savitribai Phule Pune University
## Fourth Year of Artificial Intelligence and Data Science (2020 Course)
## 417526: Computer Laboratory II: Information Retrieval

| Teaching Scheme: | Credit | Examination Scheme and Marks |
|---|---|---|
| PR: 04 Hours/Week | 02 | Term Work (TW): 50 Marks<br>Practical (PR): 25 Marks |

**Companion Course:** Elective IV: Information Retrieval (417524(B))

**Course Objectives:**

- To understand the concepts of information retrieval and web mining
- Understand information retrieval process using standards available tools

**Course Outcomes:**

**CO1:** Apply various tools and techniques for information retrieval and web mining
**CO2:** Evaluate and analyze retrieved information

**Instructions:**
Any 5 assignments from group A and 1 Mini project from group B is mandatory

## List of Assignments

### Group A

1. Write a program for pre-processing of a text document such as stop word removal, stemming.
2. Implement a program for retrieval of documents using inverted files.
3. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using the standard Heart Disease Data Set (You can use Java/Python ML library classes/API.
4. Implement e-mail spam filtering using text classification algorithm with appropriate dataset.
5. Implement Agglomerative hierarchical clustering algorithm using appropriate dataset.
6. Implement Page Rank Algorithm. (Use python or beautiful soup for implementation).
7. Build the web crawler to pull product information and links from an e-commerce website.

### Group B

Develop any one of following Mini Project-

1. Develop Document summarization system.
2. Develop Tweet sentiment analysis system.
3. Develop Fake news detection system.

## Learning Resources

**Text Books:**

1. C. Manning, P. Raghavan, and H. Schütze, ―Introduction to Information Retrievall, Cambridge University Press, 2008, -13: 9780521865715
2. Ricardo Baeza-Yates, Berthier Riberio–Neto, Modern Information Retrieval, Pearson Education, ISBN: 81-297-0274-6
3. C.J. Rijsbergen, Information Retrieval, (www.dcs.gla.ac.uk), Second Edition, ISBN: 978-408709293
4. Ryan Mitchell, Web Scraping with Python, O'reilly

**Reference Books:**

1. S. Buttcher, C. Clarke and G. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines" MIT Press, 2010, ISBN: 0-408-70929-4
2. Amy N. Langville and Carl D. Meyer, "Google's PageRank and Beyond: The Science of Search Engine Rankings", Princeton University Press, ISBN: 9781400830329

**e-Books:**

1. http://nlp-iiith.vlabs.ac.in/

## The CO-PO Mapping Matrix

| CO/ PO | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CO1 | 1 | 1 | 2 | 3 | 2 | - | - | - | - | - | - | 1 |
| CO2 | 1 | 1 | 2 | 3 | 2 | - | - | - | - | - | - | 1 |