# IT458 – Information Retrieval

## Course Overview

# Basic information

▸ Instructor:  Sowmya Kamath S.

▸ E-mail: sowmyakamath@nitk.edu.in

▸ Course Details:
   ▸ Course Code: IT458
   ▸ L-T-P:         (3-0-2)4

# Topics of Coverage

▸ **Introduction**: Fundamentals of Information Processing, basic concepts of Information Retrieval, IR system architecture.

▸ **Basic IR Models**: Text preprocessing techniques, BoW, Boolean model, inverted indexing; Vector-space retrieval models; ranked retrieval; Term frequency/inverse document frequency weighting; Fuzzy Set Models, Probabilistic models, Best-ranking models, Structured IR models

▸ **Experimental Evaluation of IR**: Performance metrics: Set based, Range based, Averaging based, Mean based, Grading based, Rank correlation, User-oriented, Evaluations on benchmark text collections.

▸ **Relevance Feedback:** Implicit and Explicit feedback, RF for VSM, Rocchio's Algorithm, Probabilistic models, Query expansion and reformulation Techniques.

▸ **Web IR:** Search engine architecture; spiders and crawlers; link analysis algorithms – HITS, PageRank, Personalized PageRank, Learning to rank.

▸ **Recommender Systems**: Content based, collaborative filtering, weighting based, Hybrid recommenders, open issues and challenges.

▸ **Challenges and Open Issues,** Domain-specific Applications and Research directions.

# Reference Textbooks

1. Manning, Raghavan and Schütze, *Introduction to Information Retrieval*, Cambridge Press. 2008.

2. C. Zhai and Sean Massung. 2016. *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. ACM and Morgan & Claypool

3. Richardo & Bertheir, *Modern Information Retrieval*, Pearson Education, 2000

4. Korfhage Robert R, *Information Storage and Retrieval*, John Wiley & Sons, Inc, 1997.

*\* Relevant IEEE/ACM/Journal papers will also be used as additional references.*

# Grading Policy

The course grade will be decided on the following criteria –

- ▶ Quizzes/Class test - 5%

- ▶ Lab Assignments  - 15%

- ▶ Mini Project  - 20%

- ▶ Mid Sem  - 20%

- ▶ End Sem  - 40%

**Homework and Assignments**

Selected topics may be set as homework/assignments whenever background study is required.

# Group Project Component.

▸ Team size – 3/4 member groups

▸ Project Requirement:

  ▸ Select a relevant problem statement/IEEE/ACM/Journal paper in the area of Information Retrieval and Management.

  ▸ Individual contribution of each student will be the major factor considered.

  ▸ Final approval will be given after discussion with each group.

# Group Project Component.

▶ Team size – 3/4 member groups

▶ Project Requirement:

    ▶ Select a relevant problem statement/IEEE/ACM/Journal paper in the area of Information Retrieval and Management.

    ▶ Individual contribution of each student will be the major factor considered.

    ▶ Final approval will be given after discussion with each group.

▶ *NOTE: Some relevant problem statements will be provided, you may check these to find those that interest you.*

# Team Project Component

▶ Final marks will be based on the following grading criteria:

  ▶ Midsem Evaluation

  ▶ Endsem Evaluation (mapped to individual student's work)
    ▶ **Part 1:** Level of work completion and implementation/coding effort
    ▶ **Part 2:** Presentation and Technical Report
    ▶ **Part 3:** Level of performance evaluation, results and analysis
    ▶ **Part 4:** Individual contribution *(for each student's overall work)*
    ▶ **Part 5:** Exceptional work *(Only given for work that results in publications/good quality software/participation in shared tasks etc)*

      \* purely on Instructor's discretion.

# Project Milestones

▸ **By Aug 12 :** Form a group and submit team details to CR (members' names and reg-nos).

▸ **Weeks of Aug 12-28:** Selection and discussion with Instructor/Mentors regarding team project.

▸ **Aug 29** – Submission of approved project (CR to maintain a spreadsheet)

▸ **Week of Sep 26–30:** Midsem evaluation.

▸ **Week of Nov 7–10:** Endsem evaluation and viva-voce.