

CS4051 - Information Retrieval

Course Outline

Instructor Waqas Ali ✓
Email waqas.ali@nu.edu.pk ✓
Semester Spring 2022
Batch BSCS 18

Note

This is a rough outline of the course and may be subject to change in the future.

Course Description

Information Retrieval is the process by which a computer system responds to a user's request for text-based information on a certain topic. IR was one of the first and is now one of the most pressing issues in natural language processing (NLP). We will cover basic and advanced strategies for developing text-based information systems in this course, which will include the following topics:

- Efficient text indexing
- Boolean and vector-space retrieval models
- Evaluation and interface issues
- IR techniques for the web, including crawling, link-based algorithms, and metadata usage
- Document clustering and classification
- Traditional and machine learning-based ranking approaches

Textbooks

1. Introduction to Information Retrieval. Manning, C.; Raghavan, P.; Sch  tze, H. Cambridge University Press.
2. Modern Information Retrieval: The Concepts and Technology behind Search. Baeza-Yates Ricardo and Berthier Ribeiro-Neto. Addison-Wesley.

Grade Distribution

| | |
|---------------------|-----|
| Class Participation | 2% |
| Assignments | 10% |
| Quizzes | 3% |
| Project | 5% |
| Sessional Exams | 30% |
| Final Exam | 50% |

Course Policies

- **General: No makeup** quizzes or exams will be given.
- **Assignments**
 - Students are expected to work independently. **Offering** and **accepting** solutions from others is an act of **plagiarism**, which is a serious offense and **all involved parties will be penalized according to the Academic Honesty Policy**. Discussion among students is encouraged, but when in doubt, ask me.
 - **Late assignments will lose marks every day.** Students must hand in every assignment.
- **Attendance and Absences**
 - Students are responsible for all missed work, regardless of the reason for absence.

Tentative Course Outline

| Week No. | Topic(s) |
|----------|--|
| 1. | Introduction to Information Retrieval (Basic Concepts, Applications) |
| 2. | Inverted Indices: Dictionary and postings lists, boolean querying |
| 3. | Index Construction and Algorithms for postings list compression |
| 4. | Spelling correction, Scoring, term weighting and the vector space model |
| 5. | Probabilistic IR: the binary independence model, BM25, BM25F |
| 6. | Evaluation methods & NDCG |
| 7. | Systems issues in efficient retrieval and scoring |
| 8. | Classification and clustering in vector spaces (Naive Bayes, kNN, decision boundaries) |
| 9. | Text Classification |
| 10. | Distributed word representations for IR |
| 11. | Ranking |
| 12. | Link analysis |
| 13. | Crawling and near-duplicate pages |
| 14. | Personalization |
| 15. | Course Review/Future Directions |