

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE  
KATHMANDU UNIVERSITY**

**Subject: Information Retrieval**

**Course Code: AICL 316**

**Credit: 3**

**E.M.: 100**

**Type: Elective II [Theory +Practical (Lab)]**

**Course Description:**

This course introduces the principles and practices of information retrieval (IR), focusing on theoretical foundations and practical applications. Students will learn how to design, implement, and evaluate information retrieval systems, emphasizing algorithms, data structures, and performance metrics. The course will also explore integrating AI techniques in IR, such as machine learning for ranking and natural language processing for query understanding.

**Objectives:**

1. Understand the fundamental concepts and techniques of information retrieval.
2. Develop skills to design and implement IR systems.
3. Apply machine learning techniques to enhance IR systems.
4. Evaluate IR systems using standard metrics.
5. Explore advanced topics in IR, including web search, recommendation systems, and IR in social media.

**Prerequisites:**

- Data Structures and Algorithms
- Probability and Statistics
- Introduction to Artificial Intelligence

**Course Evaluation:**

Internal Examination: 50%

Final Examination: 50%

**Contents:**

1. Introduction to Information Retrieval [4 Hours]
  - 1.1. Overview of IR and its applications
  - 1.2. History and evolution of IR
  - 1.3. Components of IR systems
  - 1.4. Ethical Considerations in IR (privacy, bias, and fairness)

2. Text Processing and Indexing [5 Hours]
  - 2.1. Text preprocessing (tokenization, stemming, lemmatization)
  - 2.2. Inverted index and other indexing structures
  - 2.3. Compression techniques for IR
3. Boolean Retrieval and Vector Space Model [6 Hours]
  - 3.1. Boolean retrieval model
  - 3.2. Term frequency and inverse document frequency (TF-IDF)
  - 3.3. Vector embeddings
  - 3.4. Cosine similarity and document ranking
4. Evaluation of IR Systems [3 Hours]
  - 4.1. Precision, recall, and F1-score
  - 4.2. Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (nDCG)
  - 4.3. Test collections and evaluation benchmarks
5. Probabilistic and Language Models [4 Hours]
  - 5.1. Probability ranking principle
  - 5.2. Binary independence model
  - 5.3. Language models for IR
6. Machine Learning for IR [5 Hours]
  - 6.1. Supervised learning for ranking (Learning to Rank)
  - 6.2. Feature engineering for IR
  - 6.3. Evaluation of machine learning models in IR
7. Relevance Feedback and Query Expansion [5 Hours]
  - 7.1. Relevance feedback mechanisms
  - 7.2. Pseudo-relevance feedback
  - 7.3. Query expansion techniques
8. Web Search and Crawling [5 Hours]
  - 8.1. Architecture of web search engines
  - 8.2. Web crawling and indexing
  - 8.3. Link Analysis (PageRank, HITS)
9. Advanced Topics in IR [4 Hours]
  - 9.1. Cross-lingual and multilingual IR
  - 9.2. Image and Multimedia Retrieval
  - 9.3. Generative Retrieval: Retrieval Augmented Generation (RAG) with LLM

## 10. Case Studies and Applications

[4 Hours]

### **Textbooks:**

1. "Introduction to Information Retrieval" by Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze
2. "Modern Information Retrieval: The Concepts and Technology Behind Search" by Ricardo Baeza-Yates and Berthier Ribeiro-Neto
3. "Learning to Rank for Information Retrieval" by Tie-Yan Liu
4. "Information Extraction: Algorithms and Prospects in a Retrieval Context" by Marie-Francois Moens.

### **Additional Resources**

1. Research papers and articles from conferences such as SIGIR, WWW, and CIKM
2. Online tutorials and lectures on IR and machine learning
3. Open-source IR toolkits and libraries (e.g., Apache Lucene, Elasticsearch)