INFORMATION RETRIEVAL

1.Thông tin về học phần (General Information)

Tên học phần (Course name): Information Retrieval

Mã học phần (Course code): INT_E14125

Số tín chỉ (Number of credits): 3

Loại học phần (Course type): Compulsory

Học phần tiên quyết (Prerequisites):

Học phần trước (Previous courses):

Học phần song hành (Parallel courses):

Các yêu cầu đối với học phần (Course requirements):

- Lecture room: Projector, black board, microphone and speaker
- Laboratory:

Giờ tín chỉ đối với các hoạt động (Teaching and Learning hours):

- Lý thuyết (Lectures): 34h
- Bài tập (Exercises): 6h
- Bài tập lớn (Projects): 4h
- Thực hành (Labs): 0h
- Tự học (Individual reading): 1h

Địa chỉ Khoa/Bộ môn phụ trách học phần (Address of the Faculty/Department in charge of the course):

- Address: Faculty of Information Technology 1 - Posts and Telecommunications Institute of Technology, Km10, Nguyen Trai Street, Ha Dong District, Hanoi.
- Phone number: (024) 33510432

2. Mục tiêu học phần (Objectives)

Về kiến thức (Knowledge):

The goal of this course is to provide learners with the fundamental concepts and methods of information retrieval, including:

- basic concepts and fundamental tasks in information retrieval
- applications in information retrieval
- methods for capturing, representing, storing, organizing, and retrieving unstructured or loosely structured information.

Kỹ năng (Skills):

The aim of this course is to equip learners with skills in:

- applying the learned knowledge to solve some information retrieval tasks
- determining the effectiveness of an information retrieval system using a known document corpus.

Thái độ, Chuyên cần (Attitude):

Students are required to attend the classes and complete exercises and assignments.

3. Tóm tắt nội dung học phần (Description)

Information retrieval systems are systems that provide the ability to search for specific data or information within a collection. This course will emphasize technologies and fundamental concepts that underlie all information retrieval systems. The course consists of five chapters as follows. Chapter 1 presents the overview of information retrieval. Chapter 2 introduces concepts of text processing. Chapter 3 presents ranking with indexes with inverted index, index construction and compression, and query processing. Chapter 4 presents all about queries and interfaces. Chapter 5 describes different models using in information retrieval. And chapter 6 introduces evaluation methods which help to determine the effectiveness of an information retrieval system.

4. Nội dung chi tiết học phần (Outlines)

Chapter 1. Overview of Information Retrieval

1.1. What is information retrieval?

1.2. Examples of information retrieval problems

1.3. Approaches to information retrieval

1.4. Architecture of a search engine

1.5. Performance of an information retrieval system

Chapter 2. Processing Text

2.1. Text statistics
- 2.1.1. Vocabulary growth
- 2.1.2. Estimating collection and result set sizes

2.2. Document Parsing
- 2.2.1. Tokenizing
- 2.2.2. Stopping
- 2.2.3. Stemming
- 2.2.4. Phrases and N-grams
- 2.2.5. Document structure and markup

2.3. Link Analysis
- 2.3.1. Anchor text
- 2.3.2. PageRank
- 2.3.3. Link quality

2.4. Information extraction

Chapter 3. Ranking with Indexes

3.1. Abstract model of ranking

3.2. Inverted indexes
- 3.2.1. Documents
- 3.2.2. Counts
- 3.2.3. Positions
- 3.2.4. Fields and extents
- 3.2.5. Scores
- 3.2.6. Ordering

3.3. Index construction
- 3.3.1. Blocked sort-based indexing
- 3.3.2. Single-pass in-memory indexing

6.2. Standard test collections
6.3. Evaluation of unranked retrieval sets
6.4. Evaluation of ranked retrieval sets
6.5. Assessing relevance
6.6. System quality and user utility

5. Học liệu (Textbooks)

5.1. Học liệu bắt buộc (Required Textbooks)

[1]. Manning, Christopher, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2009.

5.2. Học liệu tham khảo (Optional Textbooks)

[2]. Stefan Büttcher, Charles Clarke, and Gordon Cormack, *Information retrieval: Implementing and evaluating search engines*. Mit Press, 2016.

[3]. Bruce Croft, Donald Metzler, and Trevor Strohman, *Search Engines: Information Retrieval in Practice*, Addison-Wesley, 2009.

6. Phương pháp, hình thức kiểm tra – đánh giá kết quả học tập học phần (Grading Policy)

Grading method	Percentage	Group/Individual
- Attendance	10%	Individual
- Exercises	10%	Individual
- Mid-term projects/exam	20%	Group or individual
- Final examination	60%	Individual

Trưởng Bộ môn **(Head of Department)**	**Giảng viên biên soạn** **(Lecturer)**
Ngô Xuân Bách	**Nguyễn Ngọc Điệp**