Arab Academy for Science and Technology and Maritime Transport
Computer Science Curriculum
Course Syllabus

Course Code: CS475	**Course Title:** Information Retrieval	**Classification:** E	**Coordinator's Name:** Dr. Nahla Belal	**Credit Hours:** 3
Pre-requisites: • CS212 (Data Structures and Algorithms) • BA304 (Linear Algebra)	**Co-requisites:** None	**Schedule:** Lecture: 2 hours Tutorial-Lab: 2 hours		
Office Hours:				

Course Description:
This course studies the theory, design, and implementation of text-based information systems. The Information Retrieval core components of the course include statistical characteristics of text, representation of information needs and documents, several important retrieval models (boolean, vector space, probabilistic, inference net, and language modeling), clustering algorithms, collaborative filtering, automatic text categorization, and experimental evaluation. The software architecture components include design and implementation of high-capacity text retrieval and text filtering systems. It also introduces web search including crawling, link-based algorithms, and Web metadata; text/Web clustering, classification; text mining.

Textbook:

Information Retrieval: Implementing and Evaluating Search Engines, by Stefan Buttcher, and Charles L. A. Clarke. MIT Press.

References:

Chakrabarti S., *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan-Kaufmann.

Course Objective/Course Learning Outcome:	Contribution to Program Student Outcomes:
1. Identify basic IR models.	
2. Understand basic tokenizing, indexing, and implementation of Vector-Space Retrieval.	SO1) Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
3. Use query operations and languages.	
4. Apply **experimental evaluation of IR.**	
5. Differentiate categorization algorithms: Rocchio, nearest neighbor, and naive Bayes.	

6. Use naive Bayes text classification for ad hoc retrieval.	(SO6) Apply computer science theory and software development fundamentals to produce computing-based solutions.
7. Identify clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM).	
Learn **information extraction and integration.**	
Course Outline: 1. Search Engines and Information Retrieval. 2. Software Architecture of a Search Engine 3. Crawls and Feeds 4. Processing Text	5. Ranking with Indexes 6. Queries and Interfaces 7. Retrieval Models 8. Evaluating Search Engines 9. B+ Trees 10. Hashing

Grade Distribution:

7th Week Assessment (30%)

12th Week Assessment (20%)

Year Work (10%)

Final Exam (40%)

Policies:

Attendance:
AASTMT Education and Study Regulations (available at aast.edu)

Academic Honesty:
AASTMT Education and Study Regulations (available at aast.edu)

Late Submission:
Late submissions are graded out of 75% (1 week late), 50% (2 weeks late), 25% (3 weeks late), 0% (more than 3 weeks late)