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**: Prof. Mohamed Shaheen **Lecturer: Dr.** Mohamed Magdy	**Credit Hours:** 3
Pre-requisites: • CS212 (Data Structures and Algorithms) • BA204 (Linear Algebra)	**Co-requisites:** None	**Schedule:** Lecture: 2 hours Tutorial-Lab: 2 hours		

Office Hours: (Room 305)
Sunday 12:30 a.m. -2:30 p.m.
Monday 12:30 a.m. -2:30 p.m

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:
Croft B., Metzler D., and Strohman T., *Search Engines: Information Retrieval in Practice*, Addison-Wesley.

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.	(SO6)
2. Understand basic tokenizing, indexing, and implementation of Vector-Space Retrieval.	
3. Apply experimental evaluation of IR.	(SO2)
4. Differentiate categorization algorithms: SVM, Rocchio, nearest neighbor, and naive Bayes.	
5. Learn clustering algorithms: agglomerative clustering; k-means	

6. Learn Information retrieval concepts in World Wide Web domain	(SO6)

Contribution to Program Student Outcomes:

Outcome 2: Design, implement and evaluate a computing-based solution to meet a given set of computing requirements in the context of program's discipline..

Outcome 6: Apply computer science theory and software development fundamentals to produce computing-based solutions.

Course Outline:

1. Introduction
2. What is Information Retrieval
3. Building Dictionary
4. TF-IDF
5. TF-IDF-2
6. Vector space
7. 7^{th} exam
8. Naïve Bayes-Text Classification
9. KNN-Text Classification
10. Rocchio-Text Classification
11. SVM-Text Classification
12. **12^{th}** exam
13. link analysis
14. crawling
15. revision
16. **final exam**

Grade Distribution:

7th Week Assessment (30%):
Exam (20%) + Assignments 10%

12th Week Assessment (20%):
Exam (15%) + Assignments 5%

Year Work (10%):
Project (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