



Course description:

An introductory course on web-searching. Information vs data retrieval. The architecture of a search engine. Web crawling. Processing text (tokenization, stemming, stopwords, link analysis and markup). Ranking algorithms based on indexes and links (eg. Kleinberg's HITS, Google's PAGERANK). Retrieval Models. Search engine evaluation. Case studies (e.g. Google cluster architecture).

Aims of the course:

The main objective of the course is to provide the essentials of internet protocols.

Intended Learning Outcomes: (ILOs)

Upon completing this course the student should:

- A. Be able to explain fundamental concepts related to Web searching and the architecture of search engines.
- B. Be able to identify and explain the output of search engines in the context of web searching.
- C. Be able to understand ranking and indexing algorithms and their limitations
- D. Be able to design a search engine architecture based on input design requirements
- E. Be able to effectively use high performance computing in the design of a Web search infrastructure.

Course structures:

Week	Credit Hours	ILOs	Topics	Teaching Procedure	Assessment methods
1	3	A	WebSearching : Introduction	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
2	6	A	Fundamentals of Information Retrieval	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
3,4	6	B	The retrieval process: Crawlers and crawling	Presentation methods and techniques, Sources of information	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation



				and Instructional Aids	b) Ist Exam c) 2nd Exam d) Activity file
5	6	B	Document Processing: Parsing and Tokenization (first exam)	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
6	6	C	Document Processing: Indexing	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
7	3	C	Modeling retrieval and ranking	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
8,9	3	D	Queries, Query processing, and Interfaces	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
10	3	E	Search engine evaluation (second exam)	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
11	9	E	Classification and categorization	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file
12,13		E	Case Studies: GFS Google MAPREDUCE model	Presentation methods and techniques, Sources of information and Instructional	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam

				Aids	d) Activity file
14		E	Other Topics: Social Search	Presentation methods and techniques, Sources of information and Instructional Aids	Diagnostic tests to identify the students level and areas of weakness Formal (stage) evaluation a) Class Participation b) Ist Exam c) 2nd Exam d) Activity file

References:

A. Main Textbooks:

Search Engines: Information Retrieval in Practice by B. Croft et al., Addison-Wesley, ISBN-10: 0136072240, 2010.

Assessment Methods:

Methods	Grade	Date
first Exam	20%	
Second Exam	20%	
Assignments (Reports /Quizzes/ Seminar / Tutorials)	10%	
Final Examination	50%	

