COURSE DESCRIPTION

Dept., Number	CS429	Course Title	Introduction to Information Retrieval Systems
Semester hours	3	Course Coordinator	Dr. Nazli Goharian, Clinical Associate Professor

Current Catalog Description

Overview of fundamental issues of information retrieval with theoretical foundations. The Information-retrieval techniques and theory, covering both effectiveness and run-time performance of information-retrieval systems are covered. The focus is on algorithms and heuristics used to find documents relevant to the user request and to find them fast. The course covers the architecture and components of the search engine such as parser, stemmer, index builder, and query processor. The students learn the material by building a prototype of such a search engine. Prerequisite: CS331 or CS401 and strong programming knowledge. (3-0-3) (T) (C)

Textbook

D. Grossman and O. Frieder, Information Retrieval: Algorithms and Heuristics, Second Edition 2004, Springer Publishers, ISBN 1-4020-3004-5 (paperback).

References

Course Outcomes

Students should be able to:

- Explain the information retrieval storage methods (Inverted Index and Signature Files)
- Explain retrieval models, such as Boolean model, Vector Space model, Probabilistic model, Inference Networks, and Neural Networks.
- Explain retrieval utilities such as Stemming, Relevance Feedback, N-gram, Clustering, and Thesauri, and Parsing and Token recognition.
- Design and implement a search engine prototype using the storage methods, retrieval models and utilities.
- Apply the research ideas into their experiments in building a search engine prototype

Relationship between Course Outcomes and Program Outcomes

The following Program Outcomes are supported by the above Course Outcomes:

- a. An ability to apply knowledge of computing and mathematics appropriate to the

discipline

- c. An ability to design, implement and evaluate a computer-based system, process, component, or program to meet desired needs
- d. An ability to function effectively on teams to accomplish a common goal
- f. An ability to communicate effectively with a range of audiences
- i. An ability to use current techniques, skills, and tools necessary for computing practices.
- j. An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices
- k. An ability to apply design and development principles in the construction of software systems of varying complexity

Prerequisites by Topic

Data Structures, Algorithm and Strong Object Oriented Programming.

Major Topics Covered in the Course

1. Introduction, Overview of IR	1.5 hours
2. IR Utilities: Parser/Tokenizer, phrase Recognition, Stemming, N-Grams	3 hours
3. Efficiency: Indexing - inverted index, memory based and sort inversion; Signature Files	3 hours
4. IR Strategies and Models: Boolean, Vector Space Model; Similarity Measures in Information Retrieval, Pivoted Normalizations	3 hours
5. IR Evaluation	3 hours
6. IR Strategy: Probablistic Model	3 hours
7. IR Utility: Relevance Feedback and other Query Expansions	3 hours
8. Efficiency : Compression	3 hours
9. Efficiency: Top Docs, Query Threshold	3 hours
10. Clustering	3 hours
11. IR Strategy: Language Models	3 hours
12. World Wide Web	3 hours
13. IR Utility: Passage Based Retrieval	3 hours
14. Efficiency: Duplicate Document Detection	3 hours
15. Relational Approach	1.5 hours
16. Student Presentations	3 hours
Final Exam	-
	45 hours

Assessment Plan for the Course

End of every semester Course Objective Assessments by CS department. End of semester Course Evaluations by IIT. Reviewed every Spring semester by CS Undergraduate Studies Committee for possible updates in the following Fall. Once every 4-5 years a detailed review of all materials for the course is made by the CS Undergraduate Studies Committee.

How Data in the Course is Used to Assess Program Outcomes (unless adequately covered already in the assessment discussion under Criterion 4)

See the assessment discussion under Criterion 4

For a computer science program

Estimate Curriculum Category Content (Semester hours)

Area	Core	Advanced	Area	Core	Advanced
Algorithms		1	Software design		1
Data structures		1	Concepts of programming languages		