



COURSE DESCRIPTION FORM

INSTITUTION Department of Computer Science & IT, University Of Sargodha

PROGRAM (S) TO BE EVALUATED For MS and PhD Level

A. Course Description

Course Code	CS-6442
Course Title	Information Retrieval
Credit Hours	3 CR
Prerequisites by Course(s) and Topics	None
Assessment Instruments with Weights (homework, quizzes, midterms, final, programming assignments, lab work, etc.)	Quizzes and Homework: 20% Midterm: 30%, Final Term: 50 % For PhD Term Paper: 30%
Course Coordinator	Dr. Qaiser Abbas
URL (if any)	http://www.clsp.org/qabbas/ir.html
Current Catalog Description	Not Available as per curriculum
Textbook (or Laboratory Manual for Laboratory Courses)	Introduction to Information Retrieval (IIR) by Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, Cambridge University Press; 1st Edition (July 7, 2008). ISBN-10: 0521865719
Reference Material	<ul style="list-style-type: none"> • Modern Information Retrieval: The Concepts and Technology behind Search by Ricardo Baeza-Yates and BerthierRibeiro-Neto, Addison-Wesley Professional; 2nd Edition (February 10, 2011). ISBN-10: 0321416910 • The Geometry of Information Retrieval by C. J. van Rijsbergen, Cambridge University Press (September 13, 2004). ISBN-10: 0521838053 • Managing Gigabytes: Compressing and Indexing Documents and Images by Ian H. Witten, Alistair Moffat, & Timothy C. Bell, Morgan Kaufmann; 1st Edition (May 17, 1999) • Information Retrieval: Algorithms and Heuristics by D. Grossman and O. Frieder, Springer; 2nd Edition (December 20, 2004). ISBN-10: 1402030045 • An Introduction to Search Engines and Web Navigation by Mark Levene, Wiley; 2nd Edition (October 18, 2010).

	<p>ISBN-10: 047052684X</p> <ul style="list-style-type: none"> • Multimedia information System by V. S. Subrahmanian, Satish K. Tripathi, Springer, (December 3, 2010). ISBN-10: 1441950427 • Multimedia mining A highway to intelligent multimedia documents by Chabane Djeraba, Springer; 1st Edition (November 30, 2002). ISBN-10: 1402072473 • Search Engines: Information Retrieval in Practice by Bruce Croft, Donald Metzler, & Trevor Strohman, Addison-Wesley; 1st Edition (February 16, 2009). ISBN-10: 0136072240 • Information Retrieval: Data Structures and Algorithms by William B. Frakes and Ricardo Baeza-Yates, Prentice Hall; 1st Edition (June 22, 1992). ISBN-10: 0134638379
<p>Course Goals</p>	<p>Information retrieval is the process through which a computer system can respond to a user's query for text-based information on a specific topic. IR was one of the first and remains one of the most important problems in the domain of natural language processing (NLP). Web search is the application of information retrieval techniques to the largest corpus of text anywhere -- the web -- and it is the area in which most people interact with IR systems most frequently.</p>
<p>Topics Covered in the Course, with Number of Lectures on Each Topic (assume 15-week instruction and one-hour lectures)</p>	<ul style="list-style-type: none"> • Boolean retrieval, An example information retrieval problem, A first take at building an inverted index, Processing Boolean queries, The extended Boolean model versus ranked retrieval, References and further reading (IIR Ch1, Week 1-2) • The term vocabulary and postings lists, Document delineation and character sequence decoding, obtaining the character sequence in a document, choosing a document unit, determining the vocabulary of terms, Tokenization, Dropping common terms: stop words, Normalization (equivalence classing of terms), Stemming and lemmatization, Faster postings list intersection via skip pointers, Positional postings and phrase queries, Biword indexes, Positional indexes, Combination schemes, References and further reading (IIR Ch2, Week 3-4) • Dictionaries and tolerant retrieval, Search structures for dictionaries, Wildcard queries, General wildcard queries, k-gram indexes for wildcard queries, Spelling correction, Implementing spelling correction, Forms of spelling correction, Edit distance, k-gram indexes for spelling correction, Context sensitive spelling correction, Phonetic correction, References and further reading (IIR Ch3, Week 5-6)



	<ul style="list-style-type: none"> • Index construction, Scoring, term weighting and the vector space model, Parametric and zone indexes, Hardware basics, Blocked sort-based indexing, Single-pass in-memory indexing, Distributed indexing, Dynamic indexing, Other types of indexes, References and further reading (IIR Ch4, Week 7-8) • Index compression, Statistical properties of terms in information retrieval, Heaps' law: Estimating the number of terms, Zipf's law: Modeling the distribution of terms, Dictionary compression, Dictionary as a string, Blocked storage, Postings file compression, Variable byte codes, γ codes, References and further reading (IIR Ch5, Week 9-10) • Scoring, term weighting and the vector space model, Parametric and zone indexes, Weighted zone scoring, Learning weights, The optimal weight g, Term frequency and weighting, Inverse document frequency, Tf-idf weighting, The vector space model for scoring, Dot products, Queries as vectors, Computing vector scores, Variant tf-idf functions, Sublinear tf scaling, Maximum tf normalization, Document and query weighting schemes, Pivoted normalized document length, References and further reading (IIR Ch6, Week 11-12) • Computing scores in a complete search system, Efficient scoring and ranking, Inexact top K document retrieval, Index elimination, Champion lists, Static quality scores and ordering, Impact ordering, Cluster pruning, Components of an information retrieval system, Tiered indexes, Query-term proximity, Designing parsing and scoring functions, Putting it all together, Vector space scoring and query operator interaction, References and further reading (IIR Ch7, Week 13-14) • Evaluation in information retrieval, Information retrieval system evaluation, Standard test collections, Evaluation of unranked retrieval sets, Evaluation of ranked retrieval results, Assessing relevance, Critiques and justifications of the concept of relevance, A broader perspective: System quality and user utility, System issues, User utility, Refining a deployed system, Results snippets, References and further reading (IIR Ch8, Week 15-16)
<p>Laboratory Projects/Experiments Done in the Course</p>	<p>Through Assignments</p>
<p>Programming Assignments</p>	<p>Through Assignments</p>



National Computing Education Accreditation Council
NCEAC



NCEAC . FORM . 001-D

Done in the Course				
Class Time Spent on (in credit hours)	Theory	Problem Analysis	Solution Design	Social and Ethical Issues
		24 Hrs	12 Hrs	12 hrs
Oral and Written Communications	Students are required to submit at least 3 written reports in the form of assignments. Presentations in the form of groups would be arranged in the last week if course would be covered before the 16 th week.			

Instructor Name _____ **Dr. Qaiser Abbas**

Instructor Signature _____

Date _____