

CSI 550: Information Retrieval Fall 2009



Instructor: Prof. Tomek Strzalkowski (tomek@albany.edu) SS-262

Textbooks:

- *Search Engines: Information Retrieval in Practice*. W. Bruce Croft, Donald Metzler and Trevor Strohman. Addison-Wesley, 2009, ISBN-13: 978-0136072249
- *Modern Information Retrieval*, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley/ACM Press, 1999, ISBN 0-201-39829-X – currently out of print

Additional reference texts:

1. *Natural Language Information Retrieval*, Tomek Strzalkowski (editor), Kluwer Academic Publishers, 1999.
2. *Information retrieval: data structures & algorithms*, William B. Frakes and Ricardo Baeza-Yates. Englewood Cliffs, N.J.: Prentice Hall, 1992.

- **Credits: 3**
- **Day and time: Fridays 2:45—5:35 PM**
- **Room: SS-262**

Course Description

This course will discuss theory and practice of searching and retrieval of text and bibliographic information. Topics covered include automated indexing, statistical and linguistic models, text classification, Boolean and probabilistic approaches to indexing, query formulation and output ranking, information routing and filtering, topic detection and tracking, as well as measures of retrieval effectiveness, including relevance, utility, miss/false-alarm. Techniques for enhancing retrieval effectiveness including relevance feedback, query reformulation, thesauri, concept extraction, and automated summarization. Experimental retrieval approaches from Text Retrieval Conferences (TREC) and modern Internet search engines (Google, Yahoo!, etc) as well as recent advances into automated question-answering methods will be discussed.

Course Goals

The course is aimed at graduate and advanced undergraduate students in **Computer Science, Information Science, Business and related disciplines**. The course is intended to prepare CS students to design and evaluate information retrieval systems. The course also aims to give students a broad understanding of inner workings of automated information retrieval systems, and how such systems interact with users and affect their productivity.