CS 5604  Information Storage And Retrieval  
Tue 7:00 pm – 9:45 pm  
Room 214

Instructor:  Jianping Zhang,  
Email:  jzhang@mitre.org
Web Site: http://www.nvc.cs.vt.edu/~jzhang/cs5604


References:
- Foundations of Statistical Natural Language Processing  
- Information Retrieval: Data Structure & Algorithms, Frakes, W.B. and  
  (currently not available).
- C.J. van Rijsbergen (1979) “Information Retrieval”, available on-line  
  (http://www.dcs.gla.ac.uk/Keith/Preface.html)
- Research and Reference Articles (provided in class or online).

Content  
This course studies the theory, design, and implementation text-based statistical information retrieval systems. The Information Retrieval core components of the course include statistical characteristics of text, representation of information needs and documents, important retrieval models (Boolean, vector space, probabilistic), text clustering algorithms, automatic text categorization, topic detection and tracking, and experimental evaluation. Topics covered Ch 1 to Ch 8 and Ch 13 plus supplementary materials.

Exam:  Midterm Exam and Final Exam. The exams will be open book.

Homework:  Three programming assignments and a paper presentation. These assignments must be done individually. Generally, late assignments will be penalized 10% per day (with weekends counting as one day).
Project: A course project is an important part of the course. The final project for the class can be either an implementation project or a written project. The implementation projects would involve the design and implementation of an information retrieval system (or a specific part of a system). The written projects must involve a detailed study, survey, and evaluation of one or more topics or systems related to information retrieval and filtering. Projects may be done individually or in groups of up to 3 people (depending the complexity of the project). Each group or individual must submit a specific project proposal to be approved. A list of project ideas will be provided. You may also propose your own ideas with my approval.

Grading:

The final grade will be determined based on the following components:

- Homework: 30%
- Project: 30%
- Midterm: 15%
- Final: 25%

The general grading scheme will be based on a curve.