CSCE 5200: Web Search and Information Retrieval
Spring 2015
Course Information & Syllabus

Instructor: Cornelia Caragea
Office: F228 Discovery Park
Email: ccaragea@unt.edu
WWW: http://www.cse.unt.edu/∼ccaragea/

Teaching Assistant: Sebastian Spaink (SebastianSpaink@my.unt.edu)

Lectures: Tue / Thr 11:30AM - 12:50PM, Room NTDP D207A
Office Hours: Cornelia: Tue & Thr 10:00am - 11:00pm or by appointment, F228 Discovery Park

Course Objective: The course objectives are to understand information retrieval algorithms and identify challenging problems on the Web.The course will cover both traditional and newly developed algorithms in information retrieval and Web search and their Web applications. Examples of topics include: indexing, processing, and querying textual data; basic retrieval models: boolean retrieval, the vector space model, probabilistic IR, "intelligent" IR systems; relevance feedback and query expansion; Web crawling and search; link analysis; text classification and clustering.

Course Work and Evaluation: There will be two exams for the course. Students will be evaluated based on the exams, homework assignments, and a class project. Students are encouraged to attend every lecture and to participate in class discussion.

Assignments are due by 11:59pm on the due date. Assignments may be turned in up to 3 days late, with a penalty of 10% for each day late. No credit will be given after 3 days. There will be no final exam for this class. The final is replaced by the project. The grading criterion is shown below:

Section	Weight
Homework	25%
Exams	40%
Project	30%
Class Participation	5%

Collaboration policies:

- You are encouraged to discuss the course material, concepts, and assignments, but you must write your answers independently.
- For each assignment, you are required to list students with whom you have discussed the assignment.
- Your submission should reflect your own knowledge and you should be able to reproduce the material you turn in at any time.
- Sharing answers will not be tolerated.
- Plagiarism will not be tolerated either.
- Appropriate citations for any external sources used in your work are mandatory. Never use sentences or phrases taken directly from a paper you are reviewing.

Prerequisites: Basic knowledge on probability and statistics, data structures and algorithms. Background in information retrieval is not required.

Targeted audience: *Graduate and undergraduate students from Computer Science and related areas.*

Attendance: Attendance is essential and thus is expected.

Required textbooks:

- Introduction to Information Retrieval by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze. Cambridge University Press, 2008. Online version available at: http://nlp.stanford.edu/IR-book/.

Other Recommended textbooks:

- Readings in Information Retrieval by K. Sparck Jones and P. Willett Morgan Kaufmann, 1997.
- Modern Information Retrieval by Ricardo Baeza-Yates and Berthier Ribeiro-Neto Addison-Wesley, 1999.

Topics: The tentative topics are as follows:

The term vocabulary and postings lists
Dictionaries and tolerant retrieval
Index construction
Index compression
Scoring, term weighting and the vector space model
Computing scores in a complete search system
Evaluation in information retrieval
Relevance feedback and query expansion
XML retrieval
Probabilistic information retrieval
Language models for information retrieval
Text classification and Naive Bayes
Vector space classification
Matrix decompositions and latent semantic indexing
Web search basics
Web crawling and indexes
Link analysis
Flat clustering
Hierarchical clustering

Americans with Disabilities Act: We cooperate with the Office of Disability Accommodation to make reasonable accommodations for qualified students (cf. Americans with Disabilities Act and Section 504, Rehabilitation Act) with disabilities. If you have not registered with ODA, we encourage you to do so. If you have a disability for which you require accommodation, please discuss your needs with the instructor or submit a written Accommodation Request on or before the fourth class day.