INFO I427: Search Informatics, Fall 2013

Course staff:
Prof. David Crandall
227 Informatics West
djcran "at" indiana.edu
Office hours: Wednesdays 2-3pm

AI: Jangwon Lee
leejang "at" indiana.edu
Office and office hours: Mondays 10:00-11:30am, Informatics West 223

Lectures: Tuesdays 9:30-10:45 in Informatics East 107
Labs: Thursdays 9:30-10:45 in Informatics West 109

Course content:
This is a course about the techniques and tools used to automatically crawl, parse, index, store, rank, and search Web information. By the end of the course, students will have implemented a complete (but simplified) web search engine. Specific topics will include:

- Overview of the web and search engines
- Review of Perl and Linux
- Web crawling
- Indexing
- Vector Space Model and similarity measures
- Retrieval and ranking
- Link analysis and PageRank
- Performance analysis
- Search and query interfaces; The Google and Yahoo APIs
- Search engine optimization
- Social, legal, and ethical issues of search technology

Prerequisites: either INFO I211 and I308, or CS C211 and C212, or some comparable introductory programming course sequence. Please note that this course will require a significant amount of programming. We will spend the first several classes reviewing basic programming concepts and learning the Perl language; depending on your programming background, you may need to invest a significant amount of out-of-class time in learning and practicing programming skills. This is a nice course to take if you would like more programming experience, even if you are not a confident programmer, as there will be a lot of hands-on work and the instructor and AI are very happy to help. However, you should not take the course if you simply do not like programming and do not want to do it.


In addition, a book about Perl 5 will be helpful. One good possibility: Schwartz & Christiansen, Learning Perl, 1997 (or a later version). Used paper copies are available for a few dollars online.

The following books are not required, but may be helpful for background material:


Schedule:
A schedule for the class, with links to readings, assignments, and other resources, will be available via OnCourse: http://oncourse.iu.edu/. Check under the "Wiki" link.

Course requirements and grading:

- Assignments (40% of course grade): There will be approximately 4 assignments, each of which will require writing Perl
programs. Assignments will be accepted up to 48 hours after the due date, but a 10% late penalty will be assessed; assignments received more than 48 hours late will not be accepted and will receive a failing grade.

- **Final project (20% of course grade):** Students will have the choice of two one of two options for the final project. The *programming option* will require developing a complete web search engine and an accompanying report. The *paper option* will require writing a research paper on a topic of your choice.

- **Final exam (25% of course grade)**

- **Quizzes and Labs (15% of course grade):** In lieu of a midterm exam, there will be occasional brief quizzes (which will not necessarily be announced ahead of time). No make-up quizzes will be offered, but the lowest quiz score will be dropped. The quizzes are not intended to be stressful events, but instead are designed to (a) help prepare students for the final exam, (b) reward students who attend class regularly, and (c) give the instructors feedback on topics of confusion.

**Academic Integrity Policy:**
*We take academic integrity very seriously.* You are required to abide by the Indiana University policy on academic integrity, as described in the Code of Student Rights, Responsibilities, and Conduct, as well as the Computer Science Statement on Academic Integrity (http://www.cs.indiana.edu/Academics/integrity.html). It is your responsibility to understand and follow these policies.

Briefly summarized, these policies require that the work you submit for course assignments, projects, quizzes, and exams must be entirely your own (or entirely that of your group, if groupwork is permitted). You may use the ideas of others but you must give proper credit. You may discuss assignments with other students (or students in other groups) at a high level, by for example discussing general methods or strategies to solve a problem, but you must cite the other student in your submission. We will respond to acts of academic misconduct according to university policy concerning plagiarism; sanctions for plagiarism can include a grade of F for the assignment in question and/or for the course and must include a report to the Dean of Students Office.

Students agree that by taking this course all required papers may be subject to submission for textual similarity review to Turnitin.com for the detection of plagiarism. All submitted papers will be included as source documents in the Turnitin.com reference database solely for the purpose of detecting plagiarism of such papers. Use of Turnitin.com service is subject to the Usage Policy posted on the Turnitin.com site.

**Religious Holidays:**
Indiana University respects the right of all students to observe religious holidays and will make reasonable accommodation, upon request, for such observances. Each year, instructors are provided with the dates of major religious holidays for which students may request accommodation. Students must submit written requests for accommodation in writing by the end of the second week of the semester. Instructors are expected to give students the opportunity to do appropriate make-up work that is intrinsically no more difficult than the original exam or assignment. (Source: Indiana University Academic Guide, https://www.indiana.edu/~vpfaa/academicguide/).