CS 5634: Data Management in Bioinformatics

CS 5634 is a database course specifically geared towards bioinformatics. Its goal is to cover all the traditional aspects of database management (data modeling, schema design, querying, and interface issues) in the context of biological problems. We will begin with the major schemes of data organization such as relational (with coverage of SQL), object-oriented, text, self-describing flat files, and XML. Next, we will cover customized data representation and storage schemes for graphs (popular for modeling biological pathways and networks). Emphasis will be placed on information integration (harnessing multiple biological databases to answer specific queries) and, to a lesser extent, on data mining (inferring patterns from biological databases).

Lectures  Mon, Wed 5:30pm-6:45pm, McBryde 219
Instructor  Naren Ramakrishnan, 1-8451, Torgersen 2160L
            naren@cs.vt.edu, http://www.cs.vt.edu/~naren
Office Hours  Mondays, Wednesdays 2-4pm, or walk in any time.
Listserv  cs5634_96858.listserv.vt.edu
           (yes, the name is rather long winded.)
Course Web Page  http://courses.cs.vt.edu/~cs5634

If you are unable to make the above times and need to meet with me, you can setup an alternative time via email. If you need adaptations or accommodations because of a disability (learning disability, attention deficit disorder, psychological, or physical), if you have emergency medical information to share with the instructor, or if you need special arrangements in case the building must be evacuated, please meet with the instructor ASAP.

Pre-requisite:  CS 5046 or graduate standing in CSA. Partially duplicates CS 5614.

Evaluation:  There will be four term projects, which will involve some heavy duty programming. Each project contributes 25% toward your final grade and can be done in groups of 2-3 students.

Keeping in Touch:  Besides the course website, the class listserv will be our primary means of communication. This list is automatically created from enrollment data for CS5634.

Workload:  The course moves at a very fast pace! You are expected to readup on material after every class and get ready for the projects. I can make the course interesting but you are expected to have a willingness to learn and to put in the required effort.

Books and Reading:  There are no required books. However, depending upon your background and interests, you might wish to purchase a classical book on relational databases (talk to the instructor for suggestions) and/or references for programming skills (e.g., O’Reilly books on SQL, XML, and Postgres). Publications from online journals and conferences will be periodically posted. Lecture summaries will also be posted on the course website.

Electronic Accounts and Programming:  You are expected to have accounts on the graduate lab network in McBryde Hall or some other equivalent facility. Familiarity with high-level programming in UNIX environments is a must. You are also expected to be a good coder and to have good problem decomposition skills.