

Bio-Informatics Data Management

Description: An introduction to the concepts of Data Base Management and Information Retrieval most commonly applicable to Bio-informatics database systems. The course is intended for Computer Science graduate students wishing to focus their academic careers in Bio-Informatics. The first half of the course discusses relational databases and the second half covers Information Retrieval systems.

Objectives: The central objective of this course is to provide graduate students in Computer Science wishing to concentrate their academic careers in Bio-Informatics with fundamental concepts of data base management and information retrieval.

Specific Objectives: At the end of this course the student will have attained:

1. Knowledge of the relational model and its query language
2. Ability to design relational databases
3. Ability to formulate complex queries in the SQL language
4. Experience with existing relational database systems
5. Ability to design client programs that interface with databases
6. Understanding of the most important Information Retrieval query models
7. Understanding of the inner workings of an IR system
8. Experience with modern software development tools and operating systems

Units: Three credits

Pre-Requisites: An undergraduate course in Data Structures.

Grading: Computing laboratories and homework exercises will account for about 50% of the final grade. Partial exams account for 25% and a final exam accounts for the remaining 25%. Some of the programming assignments will involve teams of students who will orally present their results.

Instructional Strategies: The course material will be mainly presented in lectures. Students will be assigned computing laboratories and homework exercises to complement the lectures and provide an active learning experience.

Topics to be covered in the course:

- The Relational Data Model.
- Essentials of Data Base Design: E/R Models, normalization.
- Data manipulation languages: relational algebra, SQL.
- Database API's: embedded SQL's, JDBC.
- Information Retrieval (IR): query models,
- Indexing: feature extraction, IR data structures
- Document ranking and document clustering
- Case studies of popular Bio-informatics systems