INFO B556 Biological Database Management

Indiana University School of Informatics and Computing, Indianapolis
Spring 2014

Section No.: 13367
Credit Hours: 3
Time: Wednesdays 6–8:40 pm
Location: WK 321
719 N Indiana Ave
Indianapolis, IN 46202
First Class: January 15th, 2014
Website: https://oncourse.iu.edu/portal/site/SP14-IN-INFO-B556-13367

Instructor: Jake Chen, Ph.D. (Minnesota) in Comp. Sci. & Eng., Associate Professor
Office Hours: Wednesdays, 5 pm–6 pm before class and 8:40-9:30pm after class.
Or by appointment in advance.
Office: WK 311, Walker Plaza Building
719 Indiana Avenue, Indianapolis, IN 46202 [map]
Phone: (317) 278-7604 (Office)
Email: jakechen@iupui.edu
Website: http://bio.informatics.iupui.edu/CV/

TA: Rishika Bisariya
Office Hours: by Appointment
Email: ribisari@umail.iu.edu

Prerequisites: None (Not an extension of any undergraduate or graduate course)

EXTENDED COURSE DESCRIPTION

The past two decades have witnessed rapid technological advances in biological data collection and acquisition. These advances in biotechnology enabled interrogation of cellular systems at various levels, leading to generation and collection of large-scale biological data (mostly in public databases) at an exponential rate. The explosion of biological data is leading to a paradigm shift in research methods in life sciences; from hypothesis-driven research to data driven research. In the last decade, biological database management software has helped researchers in managing vast biological information from complex, heterogeneous, and very high-dimensional biological datasets. Biological databases are becoming an essential component in bioinformatics studies.

In this course we will explore both basic data management concepts and current biological database software tools. Concepts to be introduced include the relational model, E-R data modeling, data collection methods, data integration methods, ontology standards, and database-driven analytics. Specifically, the following are tentative topics to be covered in biological data management contexts:

- Biological data types and databases
– Functional dependencies and normalization process
– Defining database schemas using relational model
– Querying databases using relational algebra and SQL
– Designing conceptual data models
– Indexing and performance issues in biological databases
– Data provenance and data annotation in biological databases
– Ontologies for biological data management
– High-performance database management and big data trends

Students are expected to review a variety of biological databases such as UniProt, GO, KEGG, and emerging databases in all major research areas within bioinformatics, from sequence analysis, to genomics, functional genomics, proteomics, and systems biology. In addition, by doing projects, students will learn how to implement a database and make it available through the web.

After successfully completing the course, students are expected to have a good understanding of the basic concepts, challenges, current research topics, and trends in selected topics of biological data management.

**REFERENCE BOOKS**

**Primary Book:**


**Reference Books:**
ISBN: 0-201-39814-1

COURSE FORMAT
The instruction style will be primarily based on a mixture of lectures and student presentations of various database and database publications. Some forms of special guest lectures and research project presentation may also be used. In the classroom, students will learn basic concepts and participate in in-depth classroom discussions. Outside of the classroom, students will learn database practical skills through guided reading and class projects.

To learn basic biological database management concepts and theories, students are expected to read current literatures as the primary textbook, write critical reviews, learn to practice self-developed examples from manuals provided by the reference databases, prepare presentations based on understanding from reading, participate in in-class discussions, and write reviews to enhance further understandings.

To learn how to use and develop bioinformatics database tools, students are expected to do it perform homework questions as we cover many database software tools throughout the class. Students are also expected to practice database and database application development skills through class projects.

The emphasis of in-class learning will be to introduce fundamental concepts. It is crucial that students seek and perform readings related to classroom topics before and after the class proactively.

PRINCIPLES OF GRADUATE AND PROFESSIONAL LEARNING
Learning outcomes are assessed in the following areas:
• Knowledge and skills mastery (K&S)
• Critical thinking and good judgment (CT)
• Effective communication (EC)
• Ethical behavior (EB)
GRADING CATEGORIES

A total of 100 points are available for distribution into the following tabulated main scoring subject areas:

<table>
<thead>
<tr>
<th>Scoring Subject Area</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Preparation</td>
<td>15</td>
</tr>
<tr>
<td>Database Paper Presentation</td>
<td>10</td>
</tr>
<tr>
<td>Presentation</td>
<td>6</td>
</tr>
<tr>
<td>Summary Report</td>
<td>4</td>
</tr>
<tr>
<td>Homework</td>
<td>20</td>
</tr>
<tr>
<td>Written Exam</td>
<td>20</td>
</tr>
<tr>
<td>Team Project</td>
<td>35</td>
</tr>
<tr>
<td>Pre-proposal</td>
<td>2</td>
</tr>
<tr>
<td>Work Plan</td>
<td>5</td>
</tr>
<tr>
<td>Progress Demo</td>
<td>10</td>
</tr>
<tr>
<td>Final Demo</td>
<td>8</td>
</tr>
<tr>
<td>Final Report</td>
<td>10</td>
</tr>
</tbody>
</table>

1. **In-class Participations 15 pts**
   - Each week, a student is elected to be the class monitor. The class monitor will hand out a signup sheet and record participation scores. As long as the sheet is submitted in time to the TA, the class monitor will earn 1 full pt for the week.
   - Each student coming to a class will automatically earn 0.5 pt.
   - Students who show good preparation and good constructive discussions as determined by the monitor and reviewed by the professor may get up to 1 additional pts for each class. No more than 1/2 of the class each time may get the full 1.5 pts to avoid grade inflation.

2. **Database Paper Presentation 10 pts**
   - Identify a database of interest to present at least 2 weeks in advance by consulting with the professor. The database to present must have a recent accompanying publication (published since 2011). The topic and paper must be approved by the professor. A one paragraph brief introduction why you selected this database resource, an URL to a reference paper, and the URL to the online web site should be submitted to TA no later than 8 days in advance to avoid a late penalty of 10%.
   - **(6 pts)** Each presentation/demo is 30-40min. A thorough analysis of the database content source, data collection methods, data cleaning/curation techniques, statistics of the database created, data formats and schema, and application use cases should be thoroughly presented. Whenever possible, the student should go with a live demo with meaningful case studies. The presentation file should be uploaded to the course site through TA within 24hrs of completing the class.
• **(4 pts)** Write up a post-presentation report that summarizes all the feedback of the work received and addresses all the questions raised about the presentation itself. The report is due by Thursday 11:59pm

3. **Homework Assignments 20 pts**
   • **(2pts each)** Provide at least 10 database practice and database paper reading report before each class where the database is scheduled to be presented.
   • In each homework, the student will need to write up a personal critique of the database that address following questions:
     - What biological problems are the database application trying to address?
     - How original are the solutions?
     - How complete is the data set?
     - How did the database developers curate and validate the content?
     - Come up with some sample queries, test it on the database, and show the results
     - Discuss similar databases and compare the results obtained
     - Discuss what is the limitation of the database application
     - Give an overall personal rating of the database on a scale of 1-5.
   • In the report, address any additional questions that either the professor or the presenting student raised before the database is presented.

4. **Written Exam 20 pts**
   • Will be performed towards the end of the semester to test working knowledge of biological database management concepts and application of theories.
   • Exams will primarily consist of simple concept-related quiz and several problem-solving questions. The problems in the exams may be different from the ones covered in the class or in the book. However, the basic concepts and skills required to solve these problems will be covered prior to the exam either in class or in the assigned readings. You need to study hard throughout the semester in order to excel in these exams.
   • The exam will be 2.5 hours closed book

5. **Database Project and Presentation 35 pts**
   • Throughout the semester, each team of students is expected to perform the development, testing, and use of a specific new biological database that address a new and challenging bioinformatics problem.
   • You need to form a team of 2-4 players. You will be evaluated and ultimately graded under three general principles:
     - Does the work involve application of important biological databases?
• Is the database design sound?
• Is the database content created correctly, uniquely, and/or comprehensively for the topic covered?
• Is the database application helping biology or bioinformatics users to easily query and browse the content?
• Is the writing and presentation of your work clear and detailed?

- Deliverables are the following:
  a. **Pre-proposal:** 1 page
  b. **Work Plan / Proposal:** 6-pages
  c. **Progress Demo and report:** Limit to 25min with Q&A. Report should be no more than 10 pages.
  d. **Final project presentation:** Limit to 25min with Q&A
  e. **Final report and package:** include the following components
     - Final report (up to 15-page limit)
     - Revised presentation slides
     - A signed cover letter showing what’s learned and statement indicating contribution of each participation
     - All software codes
     - The entire database content should be uploaded to a web server for evaluations. So an URL should be provided whenever possible.

**FINAL GRADES**

The final grades will be based on summed points by referring to the following grading conversion table as closely as possible:

<table>
<thead>
<tr>
<th>Grade Category</th>
<th>Grade</th>
<th>Score Requirement</th>
<th>Relative Class Rank Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outstanding achievement</td>
<td>A+</td>
<td>&gt;93</td>
<td>Top 10%</td>
</tr>
<tr>
<td>Excellent achievement</td>
<td>A</td>
<td>[88, 93)</td>
<td>Top 20%</td>
</tr>
<tr>
<td>Very good work</td>
<td>A−</td>
<td>[84, 88)</td>
<td>Top 30%</td>
</tr>
<tr>
<td>Good work</td>
<td>B+</td>
<td>[80, 84)</td>
<td>Median</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>B</td>
<td>[76, 80)</td>
<td>Lower 40%</td>
</tr>
<tr>
<td>Marginal (Pass)</td>
<td>B−</td>
<td>[72, 76)</td>
<td>Lower 25%</td>
</tr>
<tr>
<td>Unacceptable work (Must be repeated)</td>
<td>C−</td>
<td>[60, 72)</td>
<td>Lower 10%</td>
</tr>
<tr>
<td>Fail</td>
<td>F</td>
<td>0–59</td>
<td>-</td>
</tr>
</tbody>
</table>

* Students achieving the relative class rank as shown on the table are guaranteed the minimal corresponding grade, regardless of actual score obtained.
SUBMISSION POLICY

1. Due Time

<table>
<thead>
<tr>
<th>Submission Category</th>
<th>Default Due Time</th>
<th>Submission Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Participation Signup Sheet &amp; Scorecard</td>
<td>Friday at 5 pm the day immediately following class</td>
<td>Email scanned version both to TA &amp; Professor</td>
</tr>
<tr>
<td>Homework Assignments</td>
<td>At 11:59 pm each Sunday before class</td>
<td>Upload through Oncourse</td>
</tr>
<tr>
<td>Project Deliverables</td>
<td>At 11:59 pm of Due Date indicated elsewhere</td>
<td>Upload through Oncourse</td>
</tr>
<tr>
<td>Final Report &amp; Documentation</td>
<td>At 11:59 pm of Due Date indicated</td>
<td>Upload through Oncourse &amp; Bring a printout report &amp; CD</td>
</tr>
</tbody>
</table>

2. Late Submission Policy:

Within 48 hours of the deadline: 50% penalty
Beyond 48 hours of the deadline: 0 pts

IMPORTANT DATES

1/15  First day of class
1/29  Letter of intent (proposal) due
2/19  Project proposal due
4/2   Project progress demo and report due
4/16  Written Exam
5/7   Student project final presentation
MISSION STATEMENT

The Mission of IUPUI is to provide for its constituents excellence in
• Teaching and Learning;
• Research, Scholarship, and Creative Activity; and
• Civic Engagement.

With each of these core activities characterized by
• Collaboration within and across disciplines and with the community;
• A commitment to ensuring diversity; and
• Pursuit of best practices.

IUPUI’s mission is derived from and aligned with the principal components—Communities of Learning, Responsibilities of Excellence, Accountability and Best Practices—of Indiana University’s Strategic Directions Charter.

STATEMENT OF VALUES

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana’s capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community, both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to continuous improvement of its programs and services.

CODE OF CONDUCT

All students should aspire to the highest standards of academic integrity. Using another student’s work on an assignment, cheating on a test, not quoting or citing references correctly, or any other form of dishonesty or plagiarism shall result in a grade of zero on the item and possibly an F in the course. Incidences of academic misconduct shall be referred to the Department Chair and repeated violations shall result in dismissal from the program.

All students are responsible for reading, understanding, and applying the Code of Student Rights, Responsibilities and Conduct and in particular the section on academic misconduct. Refer to The Code > Responsibilities > Academic Misconduct at http://www.indiana.edu/~code/. All students must also successfully complete the Indiana University Department of Education “How to Recognize Plagiarism” Tutorial and Test. https://www.indiana.edu/~istd You must document the difference between your writing and that of others. Use quotation marks in addition to a citation, page number, and reference whenever writing someone else’s words (e.g., following the Publication Manual of the American Psychological Association). To detect plagiarism instructors apply a range of methods, including Turnitin.com.
Academic Misconduct:

1. **Cheating:** Cheating is considered to be an attempt to use or provide unauthorized assistance, materials, information, or study aids in any form and in any academic exercise or environment.
   a. A student must not use external assistance on any “in-class” or “take-home” examination, unless the instructor specifically has authorized external assistance. This prohibition includes, but is not limited to, the use of tutors, books, notes, calculators, computers, and wireless communication devices.
   b. A student must not use another person as a substitute in the taking of an examination or quiz, nor allow other persons to conduct research or to prepare work, without advanced authorization from the instructor to whom the work is being submitted.
   c. A student must not use materials from a commercial term paper company, files of papers prepared by other persons, or submit documents found on the Internet.
   d. A student must not collaborate with other persons on a particular project and submit a copy of a written report that is represented explicitly or implicitly as the student’s individual work.
   e. A student must not use any unauthorized assistance in a laboratory, at a computer terminal, or on fieldwork.
   f. A student must not steal examinations or other course materials, including but not limited to, physical copies and photographic or electronic images.
   g. A student must not submit substantial portions of the same academic work for credit or honors more than once without permission of the instructor or program to whom the work is being submitted.
   h. A student must not, without authorization, alter a grade or score in any way, nor alter answers on a returned exam or assignment for credit.

2. **Fabrication:** A student must not falsify or invent any information or data in an academic exercise including, but not limited to, records or reports, laboratory results, and citation to the sources of information.

3. **Plagiarism:** Plagiarism is defined as presenting someone else’s work, including the work of other students, as one’s own. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged, unless the information is common knowledge. What is considered “common knowledge” may differ from course to course.
   a. A student must not adopt or reproduce ideas, opinions, theories, formulas, graphics, or pictures of another person without acknowledgment.
   b. A student must give credit to the originality of others and acknowledge indebtedness whenever:
      1. directly quoting another person’s actual words, whether oral or written;
      2. using another person’s ideas, opinions, or theories;
3. paraphrasing the words, ideas, opinions, or theories of others, whether oral or written;
4. borrowing facts, statistics, or illustrative material; or
5. offering materials assembled or collected by others in the form of projects or collections without acknowledgment

4. Interference: A student must not steal, change, destroy, or impede another student’s work, nor should the student unjustly attempt, through a bribe, a promise of favors or threats, to affect any student’s grade or the evaluation of academic performance. Impeding another student’s work includes, but is not limited to, the theft, defacement, or mutilation of resources so as to deprive others of the information they contain.

5. Violation of Course Rules: A student must not violate course rules established by a department, the course syllabus, verbal or written instructions, or the course materials that are rationally related to the content of the course or to the enhancement of the learning process in the course.

6. Facilitating Academic Dishonesty: A student must not intentionally or knowingly help or attempt to help another student to commit an act of academic misconduct, nor allow another student to use his or her work or resources to commit an act of misconduct.

7. WARNING: Any academic misconduct will be treated very seriously. The work will be scored 0 pts and the incident shall be reported to the School. So please do not violate the rules!

OTHER POLICIES

1. IUPUI course policies: A number of campus policies governing IUPUI courses may be found at the following link: http://registrar.iupui.edu/course_policies.html

2. Classroom civility: To maintain an effective and inclusive learning environment, it is important to be an attentive and respectful participant in lectures, discussions, group work, and other classroom exercises. Thus, unnecessary disruptions should be avoided, such as ringing cell phones engagement in private conversations and other unrelated activities. Texting, surfing the Internet, and posting to Facebook or Twitter during class are generally not permitted. IUPUI nurtures and promotes “a campus climate that seeks, values, and cultivates diversity in all of its forms and that provides conditions necessary for all campus community members to feel welcomed, supported, included, and valued” (IUPUI Strategic Initiative 9). IUPUI prohibits “discrimination against anyone for reasons of race, color, religion, national origin, sex, sexual orientation, marital status, age, disability, or [veteran] status” (Office of Equal Opportunity). Profanity or derogatory comments about the instructor, fellow students, invited speakers or other classroom visitors, or any members of the campus community shall not be tolerated. A violation of this rule shall result in a warning and, if the offense continues, possible disciplinary action.

3. Bringing children to class: To ensure an effective learning environment, children are not permitted to attend class with their parents, guardians, or childcare providers.

4. Disabilities Policy: In compliance with the Americans with Disabilities Act (ADA), all qualified students enrolled in this course are entitled to reasonable
5. **Administrative Withdrawal:** A basic requirement of this course is that students participate in all class discussions and conscientiously complete all required course activities and/or assignments. If a student is unable to attend, participate in, or complete an assignment on time, it is the student’s responsibility to inform the instructor. If a student misses more than half of the required activities within the first 25% of the course without contacting the instructor, the student may be administratively withdrawn from this course. Administrative withdrawal may have academic, financial, and financial aid implications. Administrative withdrawal will take place after the full refund period, and a student who has been administratively withdrawn from a course is ineligible for a tuition refund. Contact the instructor with questions concerning administrative withdrawal.

6. **Incomplete:** The instructor may assign an Incomplete (I) grade only if at least 75% of the required coursework has been completed at passing quality and holding you to previously established time limits would result in unjust hardship to you. All unfinished work must be completed by the date set by the instructor. Left unchanged, an Incomplete automatically becomes an F after one year. [http://registrar.iupui.edu/incomp.html](http://registrar.iupui.edu/incomp.html)