August 28, 2014

yarowsky@jhu.edu

Johns Hopkins University
Department of Computer Science

Prof. David Yarowsky

(600.315 and 600.415)

DATABASES
Instructor: Prof. David Varowsky
TA: TBA, Special Review Sections, and by appointment.

Office Hours: Instructor - Wed 3-4, Tuesday/Thursday after class and by appointment.

Classroom: Hackerman B17

Meeting Time: Tu, Th: 3:00-4:15 PM

yarowsky@umich.edu
410-516-5372
Hackerman 324

TAs: Prof. David Varowsky

600.315/415 - DATABASES
Required Textbook:


Other Potentially Useful Textbooks:

- S. Feuerstein, B. Pribyl, D. Russel, *Oracle PL/SQL Programming*, O'Reilly

Required Textbook:
Course Requirements

- Class Participation: 5%
- Homeworks (4): 25%
- Midterm: 20%
- Final Exam: 25%
- Final Project: 25%

Homeworks will include paper-and-pencil exercises and MySQL exercises. The Final exam will be cumulative.
Lateness Policy

One homework assignment may be handed in up to 5 days late without penalty.

No other late homeworks will be accepted.

Final projects handed in late will receive a penalty of 10% for every day late.
Computer Science Academic Integrity Code

Academichonestyisrequiredinallworkyousubmittobegraded. You must solve all homework and programming assignments entirely on your own (Homeworks 4, Project). This means you must not show your program source code from sources other than your text (such as on-line resources), discuss assignment specifications with others in the class to be sure you understand what is required by the assignment. If you use fragments of source code from sources other than your text (such as on-line resources), you must put a reference to that effect in your homework submission. Falsifying program output or results is prohibited. Please see your professor.

You must abide by JHU's Ethics Code, available at http://jhunix.hcf.jhu.edu/~ethicsbd. Students who cheat will suffer a serious course grade penalty in addition to being reported to university officials. If there are any questions about what is permissible, please see your professor.
Homeworks in 600.315 will include 1 or more additional problems and the final project will include additional component(s) not required for 600.415.

- They will differ primarily in terms of assignments and grading.
- Exams will differ somewhat and will be graded on a different scale.
- Nevertheless, 600.415 should be manageable by advanced undergraduates and
  - 600.315.

600.315 vs. 600.415

- 600.315 will be share common lectures.
- Students are encouraged to enroll.
Databases (315/415, Fall) and Database Systems (316/416, Spring) are complementary courses and make a natural course sequence (see below).

- databases (315/415, Fall) and Database Systems (316/416, Spring) are

In contrast, 316/416 will focus on:

- database internals and systems, including query and join processing,
- indexing, the organization, estimation and optimization
- database internals and systems, including query and join processing,
- databases (315/415, Fall) and Database Systems (316/416, Spring) are complementary courses and make a natural course sequence (see below).

315/415 focuses on:

- how to design and use a database;
- formal database models, theory and foundations;
- database programming languages, especially SQL and PL/SQL;
- object-oriented and XML-based data models and future directions;
- implementing a database for a novel task (including practical execution including data mining and natural language interfaces);
- database programming languages, especially SQL and PL/SQL;
- formal database models, theory and foundations;
- how to design and use a database;

The final project will be application-focused (e.g., how to design an application-focused project, including practical execution including data mining and natural language interfaces).

315/415 focuses on:

- how to design and make a natural course sequence (see below).

600.315/415 VS. 600.316/416:
The course project(s) will focus on database system internals and their development.

- Database architectures, streaming and partitioning.
Can I take 316/416 as a stand-alone course without 315 or 415?
Can I take 315/316 or 415/416 as a 2-course sequence?

Yes.

The instructors will work to make this a natural 2-course sequence.

If you have not taken a prior course in databases and are interested in both the theory/applications and systems sides of the field, then this sequence makes a lot of sense and is encouraged.

However, if you have already had a prior course in databases, or intend to continue in database systems research, then you are strongly encouraged to take 316/416 and then another advanced follow-on course in database systems, transaction processing, and/or storage systems taught by Professors Ahmadi or Burns. There will be modest overlap of material (10%) but taught via different perspectives and emphases, and will serve as a good refresher.
Can I take 315/416 as a sequence?

Yes, 416 does not require 415 as a prerequisite, but you should have done well in 315 and be prepared to do some background catching up to meet the expectations of the 416 instructor.
Can I take 415/316 as a sequence?

Yes, if you are an undergraduate and would like to continue focusing on database systems and database systems internals but a less difficult level, then this sequence could make sense.
Students will be able to select final projects of interest to them from a fairly diverse set of options. They may work in teams of 1 or 2 people.

Details will be provided in class.

For most projects, students will be required to populate and test their implemented database design with substantial quantities of real world data. Implementing a full database implementation system specification and design, a project proposal will be due in early November, including a detailed final project submission, including a full database implementation in MySQL, will be due shortly after the end of classes in December.

Final Projects
Stock market news and price correlations (data mining)
Internet proxy server database
Human genome databases
Biographic database for medical robotics
Astronomical and pharmaceutical databases for research support
Representations of acoustic data for speech recognition
Fantasy hockey league
Connecticut volunteer emergency rescue organization
JHU Fencing club and Anime Club
Olympic sports data
Movie industry data (directors, producers, actors, films, etc.)
World geography and population data (from CIA World Fact Book)
Used car information (by model and year, from Edmunds)

Sample Final Project Domains (Previous Years)
Which country had the greatest number of earthquakes in 1994?

Which country had the most powerful earthquake in 1994?

List the years in which there are at least two earthquakes of magnitude greater than 7 on the same continent.

What was the average magnitude of 1996 earthquakes in Asia?

What was the magnitude of the most powerful earthquake in China in 1996?

Which country had the greatest number of earthquakes in 1996?

Sample Final Project Domains (continued)
SEGMENT 1 - SURVEY OF DATA MODELS

Object-Oriented models

- Relational query languages: SQL, QBE (Query-by-Example)
- Formal representations: Relational algebra and calculus

Relational model

- Entity-Relationship model (formal conceptual framework)

Network and Hierarchical models (of historical interest)
SEGMENT 2 - Database Design and Implementation

Formal Analysis:
- Integrity constraints
- Domain constraints
- Triggers
- Functional dependencies
- Normalization

Practical Database Implementation:
- MySQL (a detailed exploration)
- Embedded SQL (in a host language like C or Perl)
- PL/SQL and stored procedures
SEGMENT 3 - Database System Internals

- Distributed databases
- Parallel databases
- Database system architectures
- Database security
- Recovery systems
- Transaction processing
- Query optimization
- Query processing
SEGMENT 4 - Emerging Technologies and Applications

- Decision Support Systems
- Data Mining
- Data Warehousing
- Natural Language Interfaces
- Very large text databases and information retrieval
- Multimedia Databases (images, sound, video, etc.)
- XML-based data models
- DNA and Human Genome databases
- Spatial, geometric, and geographic databases
- The impact of the WWW on database technology (and vice versa.)
- 600.466 Information Retrieval and Web Agents