

ITM 692: Special Topics in Information Technology
School of Business, University at Albany, State University of New York
Part II: Database Design and Analysis Fall 2010 Syllabus

Instructor Information

Name: Sanjay Goel Email: goel@albany.edu Phone: 442-4925 Office Location: BA310b Office Hours: Monday 11:30AM – 1:00PM or by scheduled appt.	Name: Damira Pon (Guest Instructor) Email: damira@gmail.com Phone: 442-4925 Office Location: BA310b
---	--

Class Information

Time: Tuesday 2:45 – 5:35pm
Location: BA233/BA222
Dates: October 5 – December 7, 2009
Credit(s): 3
Call #: 9690
Available Lab(s): ITM (to the right of BA 233) and HRIS (to the left of BA 234) Labs

Course Websites

URL: <http://www.albany.edu/~goel/classes/fall2010/itm692>

The course website should be your main source of course material and contains all relevant course information including details on grading, projects, assignments, course schedule, etc. In addition, this should provide a “living syllabus” a will reflect any changes made to this document.

URL: <http://bls.its.albany.edu>

Student grades and announcements will be posted on Blackboard for the class. All assignments must be submitted via Blackboard to be graded.

Text and Reference Books

Text: David M. Kroenke and David Auer, Database Concepts 5/e with MYITLab, ISBN: 0132182920

Course Overview

This is an introductory course in database modeling, design, and implementation of business applications as well as data mining. It teaches the basic principles of relational database theory and use of query languages. The class emphasizes on the fundamentals of database design. The students learn to write queries in SQL and design a database using Microsoft Access. It is expected that students have already learned to use Access through a self-study module in the ITM 601 class. The data mining portion of the class covers three topics: 1) association rule mining, 2) classification, and 3) clustering. The class is very fast paced and students are expected to complete their assignments each week in order to stay with the class. The class assumes students will learn the basic interface for Microsoft Access on their own using the MYITLab that accompanies the text book.

Learning Objectives

Students will be able to:

1. Design and create entity relationship diagrams
2. Perform database normalization
3. Write queries to create and access database information using data definition language (DDL) and data manipulation language (DML)
4. Perform functions using Microsoft Access databases, reference database for the class (point & click as well as SQL interface)
5. Perform data mining via association rule mining, classification, and clustering techniques.

ASSESSMENT & GRADING

Academic Integrity Compliance: Students MUST comply with all University standards of academic integrity. As stated on the undergraduate and graduate bulletin, "**Claims of ignorance, of unintentional error, or of academic or personal pressures are not sufficient reasons for violations of academic integrity.**" If a student is discovered to NOT comply with academic integrity standards, the student will be reported to the Office of Graduate Admissions or the Dean of Undergraduate Studies Office (whichever applies) AND receive either a warning, be told to rewrite the plagiarized material, receive a lowering of a paper or project grade of at least one full grade, receive a failing grade for a project containing plagiarized material or examination in which cheating occurred, receive a lowering of course grade by one full grade or more, a failing grade for the course, or any combination of these depending on the infraction.

Examples of violations include: Giving or receiving unauthorized help before, during, or after an examination; Collaborating on projects, papers, or other academic exercises which is regarded as inappropriate by the instructor(s), Submitting substantial portions of the same work for credit more than once, without the prior explicit consent of the instructor(s) to whom the material is being (and has in the past been) submitted; misrepresenting material or fabricating information in an academic exercise or assignment; Destroying, damaging, or stealing of another's work or working materials; and presenting as one's own work, the work of another person (for example, the words, ideas, information, code, data, evidence, organizing principles, or style of presentation of someone else). This includes paraphrasing or summarizing without acknowledgment, submission of another student's work as one's own, the purchase of prepared research, papers, or assignments, and the unacknowledged use of research sources gathered by someone else. Failure to indicate accurately the extent and precise nature of one's reliance on other sources is also a form of plagiarism. The student is responsible for understanding the legitimate use of sources, the appropriate ways of acknowledging academic, scholarly, or creative indebtedness, and the consequences for violating University regulations.

If you ever have any questions about whether you could be violating academic integrity standards - ASK!

Grading Rubric

Assignments (30%): Assignments can be in-class or take-home and will normally be pair assignments. Names of all those who participated in the assignment should be listed. Assignments will consist of exercises relevant to the material discussed in class and will be provided in class and/or through the course website. Please see the Assignments section of the course site for further details and guidelines.

Project (30%): Projects should be done in groups of three (3). A different project is offered every year and may incorporate several elements from the following: creating an entity-relationship diagram, normalization, formulation of relevant queries in MS Access, analysis of the data using the data mining techniques learned, and the creation of a written project report. The project guidelines will be provided in the second class and groups will be formed. For more details and guidelines, please see the Projects/Papers section of the course site.

Exams (40%): There will be two exams in this course consisting of multiple sections (essay-style / short answer) in which you will have to apply a majority of what has been learned during the semester for assessment of individual performance. For Exam I, this can include E-R Diagram, normalization and rationalizations, creation of a data definition table, development of SQL queries based on a needs based sentence of DML, DDL, and advanced SQL queries. Exam II will include questions associated with association rule mining, classification, and clustering.

"GREAT" EXPECTATIONS

- Students are expected to complete all assignments and readings as well as attend office hours as necessary. It is important for students to inform the instructor if all available office hours interfere with other classes during the first week of class.
- Students are expected to respectfully participate in class and communicate with the instructor if there is confusion or lack of understanding of the material. In turn, the instructor will attempt to clarify any material either in-class or outside of class.
- Students can expect assignments to be graded fairly and be returned within a reasonable time period with relevant comments and for the grader to be available to discuss questions. Students are expected to set up an appointment to meet with the grader within a week of receiving an assignment, project, or exam grade to discuss any related questions.
- Students who submit late assignments, projects, or papers can expect to lose 15% off per day late from the final possible grade for the exercise unless an excuse is legitimate. The lowest in-class assignment grade will be dropped.
- If the instructor is unable to attend class or office hours due to a personal emergency, students can expect for arrangements to be made for an alternate instructor or to be informed in as a timely a manner as possible via email/phone.
- Students are expected to contact the Disabled Student Services Center and the relevant professor at the beginning of the semester and at least a week before each exam if they require additional assistance during test-taking, e.g. reading disability, problems with writing, etc.
- Students who do not show up to take an exam and do not have a verifiable legitimate excuse should expect to receive a grade of zero for that exam. Exams are expected to be closed-book unless otherwise specified by the instructor and students are expected to keep all personal electronic devices (laptops, cell phones, PDA's, etc.) stowed during exams.
- The syllabus is subject to change and students are expected to be aware of any modifications to the syllabus including, but not limited to: due dates, readings, exam dates, and project guidelines, either announced in-class and through email. The instructor is expected to get approval of the entire class prior to making any changes regarding the grading rubric.
- Students are expected to provide reliable contact information and inform the instructor of any updates to this information. Students are expected to contact the instructor via email, phone, or in person for reliable response. Blackboard will NOT be considered a reliable communication method.
- Students can expect the instructor to be open to questions and concerns, but remain impartial and fair to all students.

COURSE SCHEDULE SUMMARY

Date	Topics	Assignments
October 5	Relational Database Design	HW 1
October 12	Database Normalization	
October 19	SQL: Data Definition Language	HW 2
October 26	SQL: Data Manipulation Language	
November 2	Introduction to Data Mining and Clustering Data	
November 9	Database Exam	HW 3
November 16	Data Classification	
November 23	Association Rule Mining	
November 30	Final Project	
December 7	Data Mining Exam	

COURSE SCHEDULE

October 5, 2010 - Relational Database Design

This class will introduce students to the fundamentals of databases and concepts of relational design. Students will learn the concepts of entities, attributes, relations, and keys. Students will learn how to construct Entity-Relationship (E-R) diagrams such that they can translate abstract problems into a relational database. This class will involve significant hands-on work where students will work on several business problems for constructing E-R Diagrams.

October 12, 2010 – Normalization

This class will continue with database design using ER-Diagrams and then introduce students to the concepts of normalization where different levels of normalization will be discussed. The class will engage students in problems involving normalization.

October 19, 2010 – SQL: Basic Queries

In this class, students will learn the basic constructs of SQL language for constructing a database. Students are expected to translate their database design by creating tables, setting relationships, and populating a database.

October 26, 2010 – SQL: Advanced Queries

The students will learn constructs of the SQL language for retrieving data from the database.

November 2, 2010 – Data Mining: Introduction & Clustering

Students will learn the various techniques that fall under the rubric of data mining and understand its application in different scenarios. Consequently, clustering techniques for naturally finding patterns within data will be taught. Students will learn both hierarchical and partitioning algorithms for clustering data. They will also learn quality measures associated with clustering. These techniques will be used on data sets they receive for drawing inferences.

November 9, 2010 – Database Exam

November 16, 2010 – Data Mining: Classification

This class will cover fundamentals of classification techniques in data mining with a focus on decision trees. The difference between supervised vs. unsupervised training is described. Also the entire decision tree process is discussed including sample selection for test and training sets, construction, pruning, and different splitting techniques (i.e. information gain/entropy, gain ratio, chi-square, and Gini/population diversity).

November 23, 2010 – Data Mining: Association Rule Mining

Students will learn to find patterns within large data sets using market basket analysis for identifying association rules. For instance they will be able to determine associations between product/entree sales in grocery stores and restaurants.

November 30, 2010 – Project Class (Tentative based on class progress)

December 7, 2010 – Data Mining Exam