

ITM 692: Special Topics in Information Technology
University at Albany, State University of New York
Fall 2007 Syllabus

INSTRUCTOR INFORMATION

Name: Sanjay Goel
Email: goel@albany.edu
Phone: (518) 442-4925
Office Location: BA310b
Office Hours: Thu: 11:30-1:30pm or by appointment

CLASS INFORMATION

Time: Thu: 8:30am-11:30pm
Location: BA 233
Credit(s): 3
Call #: 6117
Available Lab(s): MIS (to the right of BA 233) and HRIS (to the left of BA 234) Labs

CLASS DATES

Duration: August 30 - December 6
Class Dates: 8/30, 9/6, 9/20, 9/27, 10/4, 10/11, 10/18, 10/25, 11/1, 11/8, 11/15, 11/29, 12/6
Off Dates: 9/13(Holiday), 11/22 (Thanksgiving)

COURSE WEBSITE: <http://www.albany.edu/~goel/classes/fall2007/itm692/>

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

WEBCT: To access WebCT, go to <http://www.albany.edu/its/webct/cmslogon.htm> and click on the “WebCT Logon” icon. Sign in using your NetID and password and click on the BITM692 course link. WebCT will be used for assignment submission and grading.

TEXT & REFERENCE BOOKS

Required: Ira Pohl and Charlie McDowell, *Java by Dissection*. Lulu.com, ISBN: 141165238X (I would advise getting the e-Book from Lulu.com for \$5 and then printing out any necessary pages. Since the authors are generous enough to offer this low-cost version, please don't engage in copyright infringement).

Recommended: Galit Shmueli, Nitin R. Patel, Peter C. Bruce, *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. Wiley. ISBN: 978-0-470-08485-4

COURSE OVERVIEW

This course provides an overview of some emerging techniques in information technology and teaches concepts of advanced programming languages. The content of the course changes from year to year as new technologies emerge. We will cover two separate topics: 1) Java programming, and 2) business intelligence (data mining). In addition, the class will feature guest lectures on current information technology-related topics.

The programming part of the class focuses on development of simple business logic in a structured programming environment. The focus is on development of logic rather than the specifics of a

programming language. Basic elements of a programming language (e.g. data types, loops, arrays, functions etc.) as well as basic concepts of object-oriented programming (e.g. abstraction, polymorphism, and inheritance) are discussed. By the end of the course, students should be able to write simple programs in the JAVA language and be able to abstract a problem into a class structure.

As society is becoming more information-driven, data mining is receiving increased attention for extracting useful information from the large amounts of data available at our disposal. The information gained can be used for a variety of applications from customer retention and increased sales, to understanding the climactic patterns in the world. This field has evolved from several existing fields including, statistics, database management, artificial intelligence, and mathematical logic. The second part of the class deals with the fundamentals of data mining with special emphasis on its application to business decision-making activities. The class discusses different data mining algorithms (e.g. clustering, network analysis, etc.) and how data mining techniques can be used to transform large quantities of data into intelligent information. The class will use the data mining tool, XLMiner (Excel) for classical data mining algorithms.

In addition, several lectures will be organized dealing with topical business and technology issues. For these talks we will invite speakers from the industry who can provide a practical perspective of different technical and business issues. The speakers will be selected and announced during the semester and interspersed with the rest of the classes based on availability.

Please be mindful that this is a fast-moving class and if you fall behind it becomes harder to dig yourself out of a hole. Let me know if you are having difficulties and I will try to help you outside of the class so that you keep up with the class.

LEARNING OBJECTIVES

Students will learn:

- Critical thinking and logic skills for problem solving
- Syntax of Java language
- Concepts of object oriented programming
- Basic tools for writing, compiling, and running programs
- Algorithms and tools for data mining
- Business practices in IT from industry experts in emerging areas

Students should be able to:

- Install the programming environment for programming in Java
- Write algorithms for simple problems
- Compile, debug, and run Java programs
- Analyze business data for finding associations in the data
- Cluster data to determine data classes
- Classify and predict data for decision making

ACADEMIC INTEGRITY

All students are expected to follow University at Albany guidelines on academic integrity (see the Academic Integrity section of the course site for more detail). Whenever you come to me with a special request, think about whether your request is unfair to the other students. I am willing to do anything to help as long as I feel it will be useful to you and I make sure that it is fair to all students in the class.

GRADING

Whenever you come to me with a special request, think about whether your request is unfair to the other students. I am willing to do anything to help as long as I am fair to all students in my classes. There will

be no make-exams or other accommodations for lateness unless there is a valid medical excuse or other similar emergency. There will be assignments in the Java and data mining portions of the class. There will also be two projects and two exams in the class, one for Java and other for data mining.

Assignments (25%): Assignments given in class are due at the beginning of the class next week and submitted through WebCT. There will be a penalty of 10% per day for late assignments. In-class and homework assignments should be done in groups of two (chosen at the beginning of the class). Assignments are typically 5-10 points each and will consist of exercises relevant to the material discussed in class. Please see the Assignments section of the course site for further details and guidelines

Projects (25%): Projects should be done in groups of four (not assigned) and will feature of JAVA programming project based on guidelines. For more details and guidelines, please see the Projects/Papers section of the course site.

Exams (40%): The exam will consist of multiple sections (essay-style and short answer) in which you will have to apply a majority of what has been learned during the semester in order to assess individual performance. Sample questions will be provided for review.

Guest Lectures/Seminars (10%): Attendance is mandatory for guest lectures and appropriate attire (business casual) should be worn. Specific assignments will be given for seminars and will be graded.

TEAMS (PAIR PROGRAMMING): This year, students will be introduced to the concept of pair programming where a pair of students works side-by-side collaborating on software development. At any given time, one student is the driver and has the control of the computer and is actively writing the code while the other student acts as an observer and a partner who is continuously monitoring the work of the other student to identify syntactic errors and algorithmic correctness. The two students switch roles periodically so that both get the experience of program solving as well as programming. The teams would thus include two students each for the assignments. For the project though four students will be allowed since the scope would be larger.

JAVA PROGRAMMING

The goal of the java portion of the class is to promote logical thinking in students while learning the syntax, semantics, and pragmatics of a programming language. The language chosen for this course is Java because of its versatility and acceptance in the software development community however any other object oriented language could be used for a similar learning experience. Even though most students in the MBA program are not expected to pursue a career in software development, it is important to learn computer programming. Learning programming not only provides you with the syntax and semantics to write instructions for a computer but develops fundamental thinking and problem solving skills. While writing a computer program the problem needs to be broken down into intricate steps, which ensures clarity of thought for the code writer. The programming process requires defining, analyzing, developing an algorithm, writing the language syntax, and debugging the program. The process of defining the problem teaches a student to articulate the problem precisely, and the processes of analyzing the problem and developing the algorithm require the writer to examine several alternative solutions that hones critical thinking skills in the user. The step of writing the code requires translating the algorithm into the syntax of the code and is perhaps the simplest of the tasks. Most students assume that the root cause of their frustrations in programming stems from a lack of familiarity with the syntax however on the contrary the fundamental problem lies in a lack of ability to think clearly. The process of debugging requires tracing through the program and identifying the root cause of an error a skill that will be required time and again in the business world to identify causes of potential failures and for examining risks of failures.

Programming Environment

To develop software in Java three things are required: 1) an *editor* to write programs, 2) a *JAVA Compiler*, and 3) an *execution environment*. There are several good editors that can be used including TextPad, WordPad, and Notepad (Windows based) and Emacs, which is Unix (or Cygwin) based system that runs on top of Windows. The advantage of TextPad is that it provides a menu bar that allows you to compile the code within the editor itself. The java compiler should be installed based on the software installation instructions that are provided to you. The compilation and execution should be done using a Command shell or Cygwin shell. Please refer to the accompanying instructions to ensure that your environment is correctly set. Some basic knowledge of UNIX will be useful in the installation of the software and setting up the environment.

DATA MINING

Data mining involves sorting through large sets of data to extract useful scientific and business information. As more and more data is collected, data mining is becoming almost an indispensable tool for managers and scientists. Data mining encapsulates a combination of techniques that are drawn from several fields, including: databases, statistics, computer science, and mathematics combined with some visualization tools. We will draw from these fields as we study different data mining algorithms and solve problems. The goal of this class is to make students understand the concepts of data mining and help them formulate business problems into structured problems that can be solved using data mining.

Development Environment

Data mining involves data storage and manipulation, data analysis, and data visualization. You should all be familiar with spreadsheets (e.g. Excel), databases (e.g., MYSQL, ACCESS), and text editors (notepad, Emacs, and Text Pad). Depending on the size of the data and the data analysis tool that is being used one or more of these can be used for data management. There are several data mining tools that can be used for data analysis. We will be using XLMiner, an Excel-based tool for mining. A 30-day trial can be downloaded at the following web address: <http://www.resample.com/xlminer/download.shtml>, however, this will not last the full duration of the data mining portion of the class and has limitations on the amount of fields it can process. A 6-month license of this tool comes bundled with the recommended data mining book.

DETAILED SCHEDULE

1. **Java Programming 1:** The first class will primarily focus on the basics of the language and the development environment. Students will work in a lab on writing simple programs. You should also become familiar with the tools that are used in the programming environment and if required, install the tools on their own machines. Specific topics of the language that will be covered are: *programming fundamentals, data types, operators, expressions, and simple IO*.
2. **Java Programming 2:** This class will deal with building logic into program using iterative loops (e.g. *for, while, and do-while*) as well as conditional statements (e.g. *if, if-else, switch*). Students will get short problems in class which they will be expected to complete during the class. In addition, the class will cover structured programming which will be done by managing complex logic through use of functions in the program. Finally, some time will be spent discussing the concept of arrays and creating arrays. Specific topics that will be covered are: *control flow and statements, functional abstraction (methods), and arrays*.
3. **Java Programming 3:** This class starts to delve into the elements of object-oriented design: *abstraction, encapsulation, inheritance, and polymorphism*. The class will discuss the structure of a Java class: *constructors, scope of variables, and interfaces*. This will be a slightly more difficult class and students are advised to do some reading/review before and after class.

4. **Java Programming 4:** Managing Input and Output (I/O) of a program is an important part of writing a program. By this time, students will already know how to read from the command line and write to the console. This class will deal more with file and stream input and output. As a part of this class, you will also learn the concepts of managing exceptions and errors during Java programming.
5. **Java Programming 5:** The fifth class will include review of the concepts learned and practice assignments. Any questions on the Java project will also be answered in this class. Depending on the level of expertise of the students and their comfort level with the class, the examination will be held in this class or in the next class.
6. **Java Programming 6:** Examination/Guest Lecture 1 (TBD)
7. **Data Mining Lecture 1:** Association rule mining is used to discover elements that co-occur frequently in a data set. This consists of multiple independent selections of elements and determining correlations between elements. For instance, if a customer purchases milk how likely is it that he also purchases eggs? Or, if a person buys a notebook, how likely is he to buy a pen? The idea is to reduce a potentially huge amount of information to parsimonious statistically supported statements. In the first class, we will focus on pattern analysis in the data to determine associations and correlations. This will primarily involve use of the apriori algorithm and its different variants for association rule mining.
8. **Data Mining Lecture 2:** Classification is a data analysis technique that helps in segregating data into discrete classes and prediction is a technique for analyzing future trends of data based on continuous functions. For instance, classification would determine whether a bank loan application is safe or risky, prediction would determine someone's buying potential in dollars given income and occupation. These two forms of data analysis can be used to extract models that to describe data classes or make future predictions. In the class, students will learn classification techniques such as *Decision Tree Classification*, *Bayesian Classification*, *Bayesian Belief Networks*, and *Rule Based Classifiers* as well as prediction techniques such as *linear regression*, *logistic regression*, and *discriminant analysis*.
9. **Data Mining Lecture 3:** Cluster analysis helps in grouping data into classes or clusters so that objects within a cluster have a high similarity to other members within a cluster, but are dissimilar to objects in a different cluster. Different attributes can be used for affinity and often distance measures are used for assessing similarity. In this class, students will learn the different types of clustering methods, such as: *partitioning methods*, *hierarchical methods*, *density-based methods*, *grid-based methods*, and *model-based methods*.
10. **Data Mining Lecture 4:** This lecture will cover some of the advanced techniques of data mining, including: *Graph Mining*, *Social Network Analysis*, and *Neural Networks*. In addition, the class will focus on the understanding of software and tools for data mining and review the previous material covered in the data mining portion of the class.
11. **Data Mining Lecture 5:** Examination/Guest Lecture 2 (TBD)
12. **Guest Lecture 3 (TBD)**
13. **Guest Lecture 4 (TBD)**