CSE5335: Web Data Management (Fall 2019)

Instructor: Leonidas Fegaras
Office Number: ERB 653 (Engineering Research Bldg)
Email Address: fegaras@cse.uta.edu
Office Hours: Tuesday and Thursday 3:30-5:00pm
Section Information: CSE 5335-001
Time and Place of Class Meetings: TuTh 2:00-3:20pm at NH 229

Description of Course Content:
This course provides an in-depth study of the area of web data management. The course primarily covers the state of the art in designing and building web applications and services, primarily focusing on issues and challenges that revolve around the management and processing of web data. The first part of this course is an intensive study of Web programming with a focus on generating dynamic, database and web-service driven web content. The second part is an in-depth study of XML technologies, focusing on issues and challenges that revolve around the management and processing of XML data. The third part is related to using data analysis tools and cloud computing to analyze large datasets.

Student Learning Outcomes: Upon successful completion of this course, students will be able to:
- use current web technologies to develop dynamic web sites
- develop web sites that use dynamic content generated from a database
- develop web services and dynamic web applications that use web services
- store data efficiently using modern distributed storage systems on the cloud
- use data analysis tools and cloud computing to analyze web data.

Prerequisites:
Prerequisites: CSE 3330/CSE 5330 (Database Systems I) or equivalent. Students are expected to have a working knowledge of Java, SQL, and basic HTML. Students without adequate preparation are at substantial risk of failing this course.

Required Textbooks and Other Course Materials:
There is no required textbook for this course but students are expected to read many online tutorials and references (links will be given out in class).

Optional Reading: Although not required, you may find the following book useful for additional background and explanation:

Descriptions of major assignments and examinations:
There will be eight small programming assignments, one midterm exam, and one comprehensive final exam.

**Attendance:**
At UTA, taking attendance is not required. Rather, each faculty member is free to develop his or her own methods of evaluating students' academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I allow students to attend class at their own discretion.

**Grading:**
The final grade will be based on
- 48% 8 small programming assignments (6% each)
- 20% midterm exam
- 32% final exam (comprehensive)

Final grades will be assigned according to the following scale:

A: score >= 90, B: 80 <= score < 90, C: 70 <= score < 80, D: 60 <= score < 70, F: score < 60,

Sometimes, lower cutoff points are used for the final grades, depending on the overall performance of the class. Students are expected to keep track of their performance throughout the semester and seek guidance from available sources (including the instructor) if their performance drops below satisfactory levels.

**Exams:**
Both exams are open notes (all notes must be securely bound in one notebook). The final exam will cover the material from the first lecture up to and including the last lecture. Once the exam grades are posted, you will have 10 business days to dispute your grade and get your exam re-evaluated. No re-evaluation will be entertained after the 10 day period. No makeup exams will be given unless there is a justifiable reason (such as illness, sickness or death in the family). If you miss an exam and you can prove that your reason is justifiable, you should arrange with the instructor to take the makeup exam within a week from the regular exam time. For any other case, you will get a zero grade for the missed exam.

**Programming Assignments:**
There will be 8 small programming assignments. Each project will be done individually. Details will be given out in class. Late project assignments will be marked 20 points off per day (out of 100 max). So, there is no point submitting a project report more than 4 days late! This penalty cannot be waived, unless there was a case of illness or other substantial impediment beyond your control, with proof in documents from the school.

**Software:**
Most projects will be done in Java (using JDK) but some will be done in JavaScript, PHP, and XQuery. Students are expected to have a working knowledge of Java, SQL, and basic HTML. The software used for the projects is open-source, free, platform-independent, and well-suited for Java. You can do most of the projects on your own PC/laptop under any platform (Linux, MAC OS X, MS Windows, etc). Directions of how to download the required software will be given out in class.

**How to do Well in this Course:**
Students who get the most out of this course will be the ones who put in the most
effort. If you want to do well, attend all the lectures, read the assigned reading material, and start early on your programming assignments. If you are having difficulty, the instructor and the GTA will be more than happy to help you. In addition to regular office hours, the best way of communication with the instructor or the GTA is through email. If you can't make it to the scheduled office hours but really need help, contact one of us for an appointment.

Drop Policy:
Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. Students will not be automatically dropped for non-attendance. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. For more information, contact the Office of Financial Aid and Scholarships (http://wweb.uta.edu/aao/fao/).

Americans with Disabilities Act:
The University of Texas at Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including the Americans with Disabilities Act (ADA). All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of that disability. Any student requiring an accommodation for this course must provide the instructor with official documentation in the form of a letter certified by the staff in the Office for Students with Disabilities, University Hall 102. Only those students who have officially documented a need for an accommodation will have their request honored. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability or by calling the Office for Students with Disabilities at (817) 272-3364.

Title IX:
The University of Texas at Arlington is committed to upholding U.S. Federal Law "Title IX" such that no member of the UT Arlington community shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity. For more information, visit www.uta.edu/titleIX.

Academic Integrity:
Students enrolled all UT Arlington courses are expected to adhere to the UT Arlington Honor Code: I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence. I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code. UT Arlington faculty members may employ the Honor Code as they see fit in their courses, including (but not limited to) having students...
acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System Regents' Rule 50101, Â§2.2, suspected violations of university's standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University.

**Electronic Communication:**
UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at [http://www.uta.edu/oit/cs/email/mavmail.php](http://www.uta.edu/oit/cs/email/mavmail.php).

**Student Feedback Survey:**
At the end of each term, students enrolled in classes categorized as "lecture," "seminar," or "laboratory" shall be directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student's feedback enters the SFS database anonymously and is aggregated with that of other students enrolled in the course. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback is required by state law; students are strongly urged to participate. For more information, visit [http://www.uta.edu/sfs](http://www.uta.edu/sfs).

**Final Review Week:**
A period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week unless specified in the class syllabus. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

**Emergency Exit Procedures:**
Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit, which is located next to the classroom. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities.

**Web Page:**
Tentative Schedule:

1. Introduction and motivation
2. Web Programming
   a. Dynamic web pages, the HTTP protocol
   b. HTML forms
   c. XHTML and CSS stylesheets
   d. Client-side programming (JavaScript)
   e. The document object model (DOM) and dynamic HTML
   f. Asynchronous server requests (AJAX)
   g. Server-side programming: PHP scripts, cookies, and sessions
   h. Servlets, Java Server Pages (JSP), Java Server Faces (JSF)
      i. Database connectivity, JDBC
   j. Web services: RESTful vs SOAP-based, WSDL
3. XML
   a. XML basics
   b. DTD and XML Schema
   c. XPath
   d. XML APIs (DOM, SAX, StAX)
   e. XSLT
   f. XQuery
4. Cloud computing
   a. Distributed file systems (HDFS, HBase, Cassandra)
   b. The Map-Reduce framework (Hadoop, Hive, Pig)
   c. Amazon Web services and Elastic Compute Cloud (EC2)
   d. Spark, Flink, Storm

Emergency Phone Numbers:
In case of an on-campus emergency, call the UT Arlington Police Department at 817-272-3003 (non-campus phone), 2-3003 (campus phone). You may also dial 911.

Last modified: 08/11/2019 by Leonidas Fegaras