

MIS 510 “Web Computing and Mining” - Spring 2012
Hsinchun Chen, Professor, Department of MIS

Instructor: Hsinchun Chen, Ph.D., Professor, Management Information Systems Dept, Eller College of management, University of Arizona

Time/Classroom: Section 1 M/W 11:00-12:15PM MCCL 134; Section 2 M/W 3:30-4:45PM MCCL 129

Instructor’s Office Hours: M/W 2:00-3:00PM or by appointment

Office/Phone: MCCL 430X, (520) 621-4153

Email/Web site: hchen@eller.arizona.edu; <http://ai.arizona.edu/hchen>

Class Web site: <http://ai.eller.arizona.edu/mis510/index.asp> (VERY IMPORTANT!)

Teaching Assistants (TAs):

- Jonathan Shan Jiang, jiangs@email.arizona.edu, MIS Ph.D. student (office: MCCL 430)
- Julian Chenhui Guo, chguo@email.arizona.edu, MIS Ph.D. student (office: MCCL 430)

TA Hours:

Jonathan: Room 433 Tu/Th 4:00-5:00PM, and Room 424 Fr 12:00-3:00PM

Julian: Room 433 Mo/We 12:30-3:00PM

Please note that the office hours for MIS510 are only available during the following days:

Feb. 1 to Feb. 15 (Mainly for programming assignment questions)

Mar. 26 to Apr. 2 (Mainly for project questions)

Apr. 16 to May. 2 (Mainly for project questions)

CLASS MATERIAL (Optional)

- IEEE Intelligent Systems Trends and Controversies (T&C) selected articles: knowledge mapping, politics 2.0, global surveillance (2009); business intelligence 2.0, opinion mining, security informatics, virtual worlds and MMOGs (2010); social intelligence, smart health, smart money (2011)
- “Data Mining: Concepts and Techniques,” Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2nd edition, 2006.
- The WEKA Data Mining Software, Ian H. Witten and Eibe Frank.
- Please consult class web site for many excellent past class project ideas and implementations. http://ai.eller.arizona.edu/mis510/syllabus/4_ProjectResources/index.asp
- Additional readings and handouts will be distributed in class and made available through the class web site. <http://ai.eller.arizona.edu/hchen/class510.htm>

OBJECTIVES OF THE COURSE

This course introduces web computing and mining techniques, systems, and applications that are suitable for developing web-based information systems in e-commerce, search engines, digital libraries, knowledge management systems, web/data/text mining, business intelligence, security informatics, and health informatics. The course contains lectures, readings, programming assignment, lab sessions, and a large hands-on group system development project. The course will cover web mining, data mining, and text mining. In web mining, we will introduce web architecture, search engines, search algorithms, web services, Web 2.0, virtual worlds, and mobile web. State of the art data and text mining algorithms are discussed in the context of modern and emerging information systems in business, engineering, security, and health informatics. Selected data mining algorithms such as neural networks, decision trees, genetic algorithms, statistical learning, and social network analysis will be presented for clustering, database segmentation, classification, open analytics, and collaboration problems. Information retrieval, natural language processing, sentiment analysis, and authorship analysis will be discussed in text mining, especially for emerging business and market intelligence applications.

Selected algorithms will be introduced in the classroom using English-like pseudo-code. Although this course requires programming, the focus of the course is on web computing and mining concepts and applications.

PREREQUISITE FOR THE COURSE

Programming experience in Java and DBMS and good Web knowledge.

GRADING POLICY

| | |
|---|------------|
| • Programming assignment (2/15) | 10% |
| • Midterm exam (3/19) | 30% |
| • Group web mining demo V1.0 (4/2) | 15% |
| • Group web mining demo V2.0 (4/25-5/2) | 30% |
| • Group web mining paper (5/9) | 5% |
| • <u>Class attendance and participation</u> | <u>10%</u> |
| TOTAL | 100% |

PROGRAMMING ASSIGNMENT (10%)

There will be one programming assignment. You will be given two weeks to complete this assignment. All student programs will be graded and returned to you in the following week. The criteria used for grading programs will be: documentation (20%), correctness (60%), and program logic and structure (20%). NO credit will be given for turning in late assignments. You will receive partial credit for a partially finished assignment. Java is recommended for the class. The programming assignment is non-trivial (500-2000 lines of code) and may take 10-60 hours of actual programming time to complete (depending on a student's programming skill).

Two TAs will be available for the class. However, they will only be available to assist in resolving Java programming questions. Each will provide 10 hours of tutoring service during programming assignment weeks. Please see them in person during official office hours. You are required to turn in a listing (hard copy) of the program along with the output(s) and/or test runs. Students may also be requested to turn in their program source code for verification. All programs need to be developed from scratch by each student without external help. Academic integrity will be strictly enforced. Consequence for cheating will be severe.

MIDTERM EXAM (30%)

The midterm exam will be closed book, closed notes and in the short-essay format (10-12 questions). The questions will be based mostly on classroom lectures. There will be NO Final Exam for this class. Academic integrity will be strictly enforced. Consequence for cheating will be severe.

GROUP WEB MINING PROJECT ("Business Web Mining Using Amazon, eBay, and Google!"): (40%)

Many e-commerce firms such as Amazon, eBay and Google are opening up their billion-dollar data troves through web Application Programming Interfaces (APIs). Thanks to the emergence of new software standards known collectively as Web Services, web sites can swap functions, features, and data with one another like never before, all in a highly automated fashion. Increasingly, e-commerce firms such as Amazon, eBay, and Google are behaving like competing software platforms (i.e., cloud computing) that can be programmed as easily as the operating system on your PC. The rise of Web 2.0, open data, and Web Services goes even further, holding out the promise of automating the links between online businesses.

For more information about web APIs of Amazon, eBay, and Google, please see:

- www.amazon.com/webservices
- developer.ebay.com
- code.google.com

Your final group project will be about “Business Web Mining Using Amazon, eBay, and Google.” Each project team will consist of 4 members of diverse skills, who will participate in the design, coding, implementation, analysis, and maintenance of a prototype web-based business. You will be using your own machine or a Hoffman E-Commerce Lab machine as the server. Students are also encouraged to use various cloud computing platforms, e.g., Amazon, Google, Salesforce.com. You are required to create a web business, with a complete web site and business functionalities for specific customers using at least one of the three main e-commerce APIs (Amazon, eBay or Google) and others. You need to present a strong business case and design attractive system features. Your project will be judged based on its system functionalities, novelty, and business feasibility. Students are required to use 10+ diverse and multimedia Web APIs (i.e., Mashups). Web 2.0 applications are strongly encouraged, e.g., YouTube, MySpace, FaceBook, Flickr, Second Life, etc. Please see programmableweb.com for many more web APIs and Mashups opportunities. Your project needs to have a “mining” or analytics component, based on contents or data provided for your web business.

Each team member will be requested to participate in all aspects of the project. At the end of the semester, each group will present their work in actual demos and presentations. Demo V1.0 will be presented (and graded, 15%) on 4/2/2012. This initial prototype should include 5+ web APIs, initial database, and web interface to the system and data. Demo V2.0 will be presented (and graded, 20%) during the last week of the semester (4/25-5/2, 2012). This version should demonstrate complete and robust system functionality, strong business case, and novelty. Each group will be rated by the instructor and other groups during their presentation, and each team member will be rated by other members in the same team. Students can consult past project for ideas; however novelty (something new) is highly valued. For 2012, students are encouraged to develop their applications on cloud platforms and with social media and/or mobile apps focus. Your final report (5%) should summarize the complete project and detail the role and actual contributions of each team member. The final report is due 5/19/2012 (the last day of the Final Exam week).

LECTURES, ATTENDANCE, AND ACADEMIC INTEGRITY

Students are required to attend all lectures on time and honor academic integrity. Missing classes will result in loss of points or administrative drop by the instructor. Students are required to send excuse notes (via email) to the instructor before missing classes. Students are permitted to bring laptop to classroom for note taking purposes, but not for checking email or web surfing. Students need to obtain prior written permission from the instructor for using laptop in classes. Professional attitude and strong work ethics are needed for this class. Students are encouraged to consult the instructor for advice and help.

LAB SESSIONS and GUEST SPEAKERS

Selected lab sessions will be provided during the semester on the following topics: Web Services, web architecture, Java, CGI, Java servlet, JSP, ODBC/JDBC, MS SQL, Mashup, Amazon, eBay, and Google APIs and hands-on exercises and experiments. Selected guest speakers will present in the class.

COURSE OUTLINE (tentative!)

| <u>DATE</u> | <u>TOPIC</u> | <u>CONTENT/NOTES</u> |
|-------------|--|-----------------------------------|
| Jan 11 | Course introduction | syllabus, roster, overview |
| Jan 16 (M) | Martin Luther King Jr. Holiday | NO CLASS |
| Jan 18 | <u>Web mining</u> overview | web computing, services |
| *Jan 23 (M) | Web services; class projects | guest lecture/lab session |
| Jan 25 | Web 1.0; Internet SE | architecture and components |
| Jan 30 (M) | Web search | graph search, spidering |
| | TEAM WEB MINING PROPOSAL DUE | |
| Feb 1 | Genetic algorithms | search algorithms |
| Feb 6 (M) | Web 2.0; social web | overview |
| Feb 8 | Social media analytics | deep web, dark web |
| Feb 13 (M) | Social media analytics | virtual worlds, MMOGs |
| Feb 15 | Web mining applications | IEEE articles, security, health |
| | PROGRAMMING ASSIGNMENT DUE (10%) | |
| Feb 20 (M) | Web structure analysis | web links and forums, SNA |
| Feb 22 | Web 3.0; SoLoMo web; cloud | mobile web, apps, VCs |
| Feb 27 (M) | <u>Data mining</u> overview | machine learning, analytics |
| Feb 29 | BI & Analytics | BI 2.0 and education |
| Mar 5 (M) | Graphs/Neural networks | classification algorithms |
| Mar 7 | LAST DAY TO DROP (with "W") | |
| Mar 8 | Classification/ID3, clustering | techniques, sports, health, crime |
| *Mar 10-18 | SPRING RECESS | NO CLASS |
| Mar 19 (M) | MIDTERM EXAM (30%) | |
| Mar 21 | SVM, statistical learning | tools, examples, applications |
| Mar 26 (M) | Social network analysis (SNA) | models, dynamics |
| Mar 28 | Open source analytics | Weka, MapReduce, Hadoop |
| Apr 2 (M) | PROJECT PRESENTATION/DEMO 1.0 (15%) | |
| Apr 4 | <u>Text mining</u> overview | NLP, Watson, Siri |
| Apr 9 (M) | Digital library/IR | DL initiatives |
| Apr 11 | Natural language processing | overview |
| Apr 16 (M) | Topic extraction and authorship | techniques |
| Apr 18 | Sentiment analysis | opinion mining, social media |
| Apr 23 (M) | Information visualization | HCI |
| Apr 25 | FINAL PROJECT PRESENTATION/DEMO 2.0 (30%) | |
| Apr 30 (M) | FINAL PROJECT PRESENTATION/DEMO 2.0 | |

May 2
May 4-10
May 9

FINAL PROJECT PRESENTATION/DEMO 2.0
FINAL EXAM WEEK NO EXAM FOR MIS 510
FINAL PROJECT PAPER DUE (5%)