

CS 0155: Data Witchcraft (Spring 2014)

Department of Computer Science, University of Pittsburgh

When: Monday & Wednesday, 11:00 am – 12:15 pm

Where: Room 6110, Sennott Square Building

Instructor: Prof. Alexandros Labrinidis

Email: labrinid@cs.pitt.edu

Web: <http://labrinidis.cs.pitt.edu/>

Office: 6105 Sennott Square

Phone: 412-624-8843

Office hours: Mondays 4:30pm - 5:30pm

Wednesdays 4:30pm - 5:30pm

and by appointment

Recitations: Friday 11:00am – 11:50am @ 6110 Sennott Square
(only on: January 17, February 7, February 28, March 28)

Undergraduate Teaching Assistant: Alec Jasen

Email: cs0155-staff@cs.pitt.edu

Office: 6506 Sennott Square

Phone: TBD

Office hours: Tuesdays 9:00am - 10:30am

Thursdays 9:00am - 10:30am

Course Description: This course is designed as a second course in computer science for non-CS majors and CS minors. CS Majors who want to follow a data management concentration can also take the course. The course will serve as an introduction to basic data management / data science technologies, typically through the use of different discipline-specific examples. It will be taught in a computer lab where the students can actively participate in parts of the lecture. The course will adopt the point of view of a user of data (i.e., who is just combining and analyzing it using tools) and not a provider of data (i.e., who would be implementing a database driven Web site), as is typically the case for related courses for CS majors.

Prerequisites: CS 0007 or CS 0008 or CS 0401 or permission of the instructor.

Class Web Page: <http://data-witchcraft.org>

All handouts and class notes will be published on the class web page. You are expected to check this page frequently (at least twice a week).

Google+ Page: We will use Google+ as an authentication mechanism for posting photos of the whiteboard from every class. You will need to provide a valid Gmail account to have access to the photos.

Textbook (optional): None currently assigned.

Reference: Numerous reference books/chapters will be given throughout the term, mostly through O'Reilly's *Safari Bookshelf* for which the University has institutional access (i.e., you will not have to buy extra books).

Course Grading:

Assignments	30%	There will be 4-6 assignments/projects , some of which will have a programming portion. All have equal weight.
Term project	10%	Due April 19th; Presentations during finals exam slot, on Tuesday, April 22nd, 12pm - 1:50pm.
Class participation	10%	For both lecture and recitations, including in-class quizzes.
Midterm Exam #1	25%	Wednesday, February 5th, 11:00am – 12:15pm (SENSQ 6110)
Midterm Exam #2	25%	Wednesday, March 19th, 11:00am – 12:15pm (SENSQ 6110)
Final Exam	0%	There is no final exam for this class

[please turn over]

Class Mailing List: All students must subscribe to the class mailing list, so that they receive time-sensitive information from the instructor and TAs. You will be automatically added to the mailing list.

Note on Email Communication:

You should send all email regarding class matters to **cs0155-staff@cs.pitt.edu**. Your email will go to the instructor, and the TA. If you have a confidential matter, then please email the instructor directly, but make sure to include the keyword `cs0155` in the subject line of your email messages. We will make every effort to respond to all email requests within one business day at the latest. **Due to spam filtering, you should always use your pitt email address when sending email and include your full name.**

Grading Policy:

Unless explicitly noted otherwise, the work in this course is to be done independently. Discussions with other students on the assignments should be limited to understanding the statement of the problems (except when assignments are to be done in groups in which case it is expected of members of the same group to work together). **Cheating in any way, including giving your work to someone else, will result in an F for the course and a report to the appropriate University authority.** Submissions that are alike in a substantive way will be considered to be cheating by ALL involved parties. Please protect yourselves by only storing your files in private directories, and by retrieving all printouts promptly.

Students are expected to abide by the Dietrich School of Arts and Sciences' Academic Integrity code of conduct, posted at <http://www.as.pitt.edu/faculty/policy/integrity.html>

All assignments must be submitted electronically. Grades can be appealed up to two weeks after they have been posted; no appeals will be considered after that time.

Late Policy: A late assignment will receive a deduction of 5 points if it is up to one day past the deadline and 15 points if it is up to two days past the deadline. Assignments that are past two days late will not be accepted.

Make-up Policy: Students are expected to be present for all exams and quizzes. Make-up exams will only be given in the event of an emergency, and only if the instructor is informed **in advance**. Failure to notify the instructor prior to missing an exam will result in a zero for the exam.

Students with Disabilities:

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services, 216 William Pitt Union, 412-648-7890 or 412-383-7355 (TTY) as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course. Their web site is <http://www.drs.pitt.edu>.

Religious Observances:

In order to accommodate the observance of religious holidays, students should inform the instructor (by email) of any such days that conflict with scheduled class activities **within the first two weeks of the term.**

Audio/Video Recording:

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

[please turn over]

Cell Phone Use:

Answering a cell phone or texting is disruptive and hence any use of a cell phone is not permitted in the class or recitation. Cell phones must be switched to silent mode and if you have a phone call which cannot wait until the end of the class, you need to step out of the class and then answer it.

Outline:

A detailed reading guide will be published on the web page, along with the class notes and additional online articles and resources. Time permitting, we will cover the following topics:

1. Introduction: Data-intensive science and the promise of Big Data
2. Introduction to Information Retrieval
3. Introduction to Data Mining
4. Introduction to Big Data Analytics
5. Introduction to Perl programming
6. XML / RSS
7. Querying XML (XPath / XQuery)
8. Popular data exchange formats (CSV, XML, RDF, KML, JSON)
9. Introduction to the Semantic Web
10. Querying RDF (SPARQL)
11. Using SQL (MySQL)
12. Using Google Fusion Tables
13. Ontologies

[Last updated on January 6, 2014 at 10:49am EST]