

Syllabus

Fall 2007

Course: CS262: Information and Knowledge Management
Time: MWRF 6th hour
Room: MWF: SMC D-205
Th: SMC E-011 (Cat Lab)
Website: <http://courses.knox.edu/cs262/>

Professor: Don Blaheta
Office: SMC E-210
Phone: x7956
Email: dblaheta@knox.edu
Office hours: Wednesdays 4–5:30pm;
Tuesdays and Thursdays noon–1pm

Overview

In the early days of computers, the field was all about number crunching. Increasingly in the last several decades, we've seen the research problems of computer science become much more about how to store, process, and interact with information. Though not every computer scientist need be an expert in areas like artificial intelligence, databases, and human-computer interaction, one should have a general grounding in all of them.

Such is the aim of this class. We will cover a variety of disparate topics in computer science that will provide you with the tools you need, whether to continue in those subfields or just to understand what your colleagues are talking about.

Book

Due to the smorgasbord nature of the course, there is no single appropriate textbook, and therefore no required textbook. I will be sourcing from several different textbooks, including significant portions of the following, which I will put on reserve in the library:

Baase. *A gift of fire: social, legal, and ethical issues for computers and the Internet.*

Garcia-Molina, Ullman, and Widom. *Database systems: the complete book*. (If you're buying, Ullman and Widom's *A first course in database systems* is identical for the chapters we're using.)

Luger. *Artificial intelligence: structures and strategies for complex problem solving*.

Manning and Schütze. *Foundations of statistical natural language processing*. Mostly for the chapters on conditional probability; also a bit for evaluation measures.

In addition, Wikipedia and other websites often have a pretty good explanation on the topics we're discussing, so I'll be posting links to the course website on a semi-regular basis.

Contacting me

I'll be around during my listed office hours for sure, but any time my door is open (which is most of the time) you should feel free to come in and ask questions or just to chat. If you can't catch me in my office, email is probably your best bet. If you don't get a quick response to email, I do respond well to (light) nagging—you might want to try sending another one.

Graded work

I figure that I have about a third of your time over the course of the term, or about 12–15 hours a week including class. This time is spread over a number of different types of graded assignments. Note that this means I expect roughly 4–5 hours outside of class for each 5% of the grade—budget your time accordingly.

Homework and projects. Homework will go out relatively frequently, and will often be short and due soon. Sometimes a “homework” will be a bit larger, worth more points, and you'll have a longer time to work on it. Both the regular homeworks and the more projecty assignments will be a mix of on-paper work (e.g. simulating an algorithm, describing concepts) and implementation problems. Feel free to discuss them amongst yourselves, but be certain to do all the writing yourself (v. the collaboration policy; homeworks and projects are what I term “lightly collaborative”).

Presentations. Roughly once a week, we'll have a more discussion-based class, which will be led off by a 15–20 minute presentation by one of you regarding some issue affecting computers or the internet. Each person will do one of these. Signup for these will be early in the term, and you'll be responsible for planning the short talk and arranging a practice talk with me.

Exams. There will be two exams, one at midterm and one for the final. Both will be take-home, and you will be given several days to work on them. These will be non-collaborative, and you are not permitted to discuss the exam with anyone other than me.

Labs and participation. A small portion of the grade is devoted to making sure you're attentive in lecture and active in labs. Occasionally asking questions, or answering them, is a good way to lock in these points. With respect to labs, you don't lose points for not finishing the lab in the allotted time, nor do you lose points for finishing early and leaving (after checking with me); mostly you just lose points if you're goofing off (or absent). So don't do that.

Breakdown

Homework/projects	45%
Presentation	10%
Midterm exam	20%
Final exam	20%
Participation	5%

Schedule

Homeworks will go out intermittently and typically be due the following class period. All assignments are due at the beginning of class on the specified due date.

The midterm exam will go out Monday, 1 October, and be due Thursday, 4 October at the start of the lab period.

The final exam will be handed out on the last day of class (Monday, 12 November), and it will be due at the beginning of the course's official exam period. You are, of course, permitted to hand it in early, either in person or under my office door.

Topics

Topics in [brackets] will be the subject of student presentations and subsequent class discussion.

Week 0

Propositional & predicate logic; Prolog. Horn clauses.

Week 1

Theorem proving; default logic; nonmonotonic inference. Probabilistic reasoning. Conditional and joint probability.

Week 2

Bayes' theorem and Bayesian inference. [Overview of information as property (IP): copyright, patents, trade secrets, software patents, the public domain.] Corpus-based information processing. Information retrieval; precision and recall. Bandwidth and latency.

Week 3

Information storage; data independence; data files vs. databases. Database components; DB correctness properties (ACID). [Information as value: from copyright to customer databases and Google.] Lossless and lossy compression. Huffman coding, Lempel-Ziv compression, DEFLATE.

Week 4

[Violating copyright: Early history, software "piracy", filesharing.] [What is intelligence? The Turing test, Eliza, and the Chinese room.] Storing world knowledge; seeking optimality vs. emulating humans (reasoning, behaviour). Problem spaces. Brute-force search: depth-first, breadth-first, iterative deepening.

Week 5

Heuristics. Best-first search: Dijkstra's algorithm, A and A* (and admissibility). Minimax, alpha-beta pruning, backtracking. [Free speech and censorship online.]

Week 6

Data presentation; interacting with humans; perception and cognition (affordances, conceptual models, feedback); culture and communication; accommodating diversity. Principles of HCI and associated tradeoffs; usability testing. UI design principles.

Week 7

[Open source software, free software.] Modelling information: Entity-relationship model, Frame-based AI. DB design principles. Constraints. Object-oriented models; relational models. Relational design flaws; SQL.

Week 8

Converting to relational schemata: decomposition; functional dependencies. Relational algebra. Web transactions; CGI. Web-related database issues. Privacy, integrity, security, preservation; [Privacy: ethics, civil liberties, implications of massive databases.]

Week 9

[History of cryptography: early forms, WWII, online.] private key crypto and the key exchange problem. Public key cryptography; digital signatures. [Online crime.]

Week 10

[Computer ethics: Granting and exercising access to information.]

Presentation dates

These are subject to a little bit of float, but in general, you should plan to give your presentation on the date given below, and schedule a practice talk with me for the day before. I won't move them earlier without at least a week's notice.

- Wednesday, 19 September: IP Overview
- Wednesday, 26 September: Information as value
- Monday, 1 October: Violating copyright

- Wednesday, 3 October: What is intelligence? (NOTE: practice talk no later than 30 September)
- Friday, 12 October: Free speech.
- Monday, 22 October: Free software.
- Friday, 2 November: Privacy.
- Monday, 5 November: Cryptography.
- Friday, 9 November: Online crime.
- Monday, 12 November: Information ethics.

Policies

The “read the damn spec already” policy

It is vitally important that you, as a budding computer scientist, learn to write programs according to spec. In addition, although I do give partial credit, my grading task often becomes infeasible if I have to fix your code before I can even test it. Therefore, on programming problems (which you will hand in electronically), I reserve the right to give you **ZERO** points if it:

- doesn’t compile,
- doesn’t accept the right command line,
- doesn’t open the right file, or read the right thing from the keyboard,
- doesn’t output to the right place (file, screen, window, whatever), or
- doesn’t follow the given API.

If you do one of these things, you might get lucky and get partial credit. Or, you might not. {Insert Clint Eastwood reference here.} In general, a program that compiles, runs, and does only a little bit will get more credit than a program where “all the parts are there” but it won’t even compile.

I’ll generally try to verify at least that handed-in work passes this test soon after the due date. If I find a problem, I’ll give it back to you, and you’ll

be able to fix it and resubmit. You won't be changing other parts of the problem; just the interface (I'll check). And when you hand it back in, it will be graded at a penalty of 10% of the maximum points for the problem—thus, an automatic -1 on a ten point problem.

Collaboration policy

Exams are noncollaborative. The other assignments (homeworks, projects) are “lightly collaborative”. What I mean by this can be found at <http://faculty.knox.edu/dblaheta/collab.html> .

Systems and environments

The supported systems in this class will be the department Linux boxen in the Cat lab. You are welcome to use any systems you like for development, but the work you hand in has to be able to run on those machines. Also, I can't guarantee that I'll be able to help as much on other systems.

Attendance and late policy

Attendance is required, and assignments must be turned in on time. That said, if you have a good reason to miss class or hand something in late, I tend to be fairly easygoing about it; but I need to hear about it in advance. (Medical and family emergencies excepted, of course. But get a note from a dean.) Extensions of a day or two will be granted fairly routinely, as long as the reason is something better than “I didn't start until today, and now I can't finish on time!” If in doubt, come talk to me. Frequent absence will result in a lowered participation grade; habitual absence may in extreme cases result in a failing grade for the class. Unexcused late assignments will normally be given a zero.

Early bird policy

My standard early bird policy applies: extra credit if you find an error in the problem definition (for homework, project, exam, whatever). See <http://faculty.knox.edu/dblaheta/policies.html#early> for details.