

SYLLABUS

LING 388 LANGUAGE & COMPUTERS

TIME: Tuesdays and Thursdays 11:00 AM - 12:15 PM

PLACE: S SCI 224

INSTRUCTOR: Jeff Berry

OFFICE: Douglass 304

OFFICE HOURS: TBA, by appointment

EMAIL: jjberry@email.arizona.edu

1. GOALS OF THE COURSE

This course is a hands-on introduction to commonly used tools in natural language processing and computational linguistics. We will focus on learning how to use regular expressions, finite state machines, and context-free grammars to solve real-world problems. We will build on previous knowledge as the course progresses. We will apply our knowledge to constructing a dialogue manager to handle text returned from an automatic speech recognizer. By the end of the course, you will have basic proficiency with the following software:

- bash for scripting and programming
- grep, sed and awk for text processing with regular expressions
- foma: finite state machine toolkit
- flex and bison for constructing parsers
- julius for automatic speech recognition

All of the software we will use is open-source, meaning you don't have to pay for it. It is available on all major platforms including Linux, Mac, & Windows (through Cygwin).

2. REQUIREMENTS

Due to the hands-on nature of the course, attendance is mandatory. Valid, documented absences will be accepted, with the expectation that you will make arrangements with me to cover the missed material. 29% of your final grade will be based on attendance and participation (1% for each lecture). These points will be awarded based on submission of short in class exercises that we will work through as a group. There will be a short quiz each Thursday at the beginning of class that will require implementation of a concept discussed in class. There will be 4 homework assignments due roughly every third week (see schedule for details). The homework assignments are designed to build the necessary tools for the final project: a dialogue manager for a speech recognizer. All quizzes and homeworks should be submitted as an executable program with instructions for me to run on my computer. No late work will be accepted unless you have a valid, documented excuse, or make arrangements with me previously. Extra credit will be awarded for participation in linguistics experiments.

Date: Fall 2009.

TABLE 1. Breakdown of requirements

Class participation/attendance	29%
Weekly quizzes	11%
Homework 1	10%
Homework 2	10%
Homework 3	10%
Homework 4	10%
Final project	20%

3. TEXTBOOKS & READINGS

This course has no required textbook, however you are encouraged to purchase a copy of the following books, for future reference. Scans of the chapters we will be covering in class will be provided, but we will only be scratching the surface of the amount of useful information these books have to offer.

- Kernigan, Brian & Rob Pike. (1984). *The UNIX programming environment*. Prentice Hall.
- Friedl, Jeffrey. (2006). *Mastering regular expressions, 3rd edition*. O'reilly Media.
- Dougherty, Dale & Arnold Robbins. (1997). *sed & awk, 2nd edition*. O'reilly Media.
- Beesley, Kenneth & Lauri Karttunen. (2003). *Finite state morphology*. CSLI.
- Levine, John. (2009). *flex & bison*. O'reilly Media.

Other readings may be assigned and will be available on the course website.

4. TIPS FOR SUCCESS

The technical material in this course may be difficult for students who have no prior programming experience. Learning to think like a computer can be difficult and occasionally frustrating. If you are struggling to understand the material, please ask questions in class, make use of office hours, and talk to your peers. Most importantly, don't get discouraged!

5. UNIVERSITY POLICIES

- All holidays or special events observed by organized religions will be honored for those students who show affiliation with that particular religion. If this is relevant for you, you must inform the instructor during the first two weeks of class.
- Absences pre-approved by the UA Dean of Students (or Deans designee) will be honored. If this is relevant for you, you must inform the instructor during the first two weeks of class.
- Cell phone policy: cell phones should be off; they are not permitted during class.
- Plagiarism is not acceptable. See Student Code of Academic Integrity.
- Threatening behavior is not allowed.
- It is possible that some course content may be deemed offensive by some students.
- Students who are registered with the Disability Resource Center must submit appropriate documentation to the instructor if they are requesting reasonable accommodations. If this is relevant for you, you must inform the instructor during the first two weeks of class.

TABLE 2. Tentative course schedule (subject to change)

	Date	Topic	Readings	Due dates
1	08/25	Introduction		
2	08/27	Recursion in natural lang		
3	09/01	Chomsky hierarchy		
4	09/03	Unix basics	Kernighan & Pike selections	
5	09/08	grep: regular expressions		
6	09/10	egrep: extended regular expressions	Friedl Ch 1	
	09/15			
7	09/17	guest lecture: careers in comp ling		HW 1
8	09/22	guest lecture		
9	09/24	sed: regular expressions in text processing	Dougherty & Robbins Ch 2	
10	09/29	awk: regex programming		
11	10/01	more awk	Dougherty & Robbins Ch 7-9	
12	10/06	regex wrap-up		
13	10/08	FSA intro: equivalence to regex		HW 2
14	10/13	foma: finite state machine toolkit	Beesley & Karttunen Ch 1-2	
15	10/15	foma: finite state phonology		
16	10/20	foma: finite state morphology	Beesley & Karttunen Ch 3 sections	
17	10/22	foma: finite state morphology		
18	10/27	Markov chains: adding probability to FSA		
19	10/29	Context-free grammars vs. natural lang		HW 3
20	11/03	Context-sensitive grammars		
21	11/05	flex: lexical analyzer / bison: parser		
22	11/10	more flex/bison: building an English grammar	Levine Ch 1-3 sections	
23	11/12	more flex/bison: building an English grammar		
24	11/17	more flex/bison: building an English grammar		
25	11/19	building a grammar for an ASR system		HW 4
26	11/24	building a dialogue manager		
	11/26	Thanksgiving		
27	12/01	putting it all together		
28	12/03	help with final project		
29	12/08	Wrap-up: a look back, next steps		
	12/17			Final