

Syllabus

INLS 512, Applications of Natural Language Processing, Spring 2014

Instructor: Stephanie W. Haas
Phone: 919-962-8360

Office: 111 Manning Hall
Email: shaas at email dot unc dot edu

Office Hours: Monday and Wednesday 11:00 – 12:00 and by appointment
Class Meeting: 014 Manning Hall, Monday and Wednesday 9:30 – 10:45

Syllabus Contents

- Course Description
- Rationale and Approach
- Keys to Success
- Learning Objectives
- Course Policies
- Policies on Academic Integrity and Diversity
- Grading Policies
- Communications
- Sakai

Course Description

INLS 512, Applications of Natural Language Processing

Applications of natural language processing techniques and the representations and processes needed to support them. Topics include interfaces, text retrieval, machine translation, speech processing, and text generation.

This course is open to undergraduate and graduate students who have an interest in the intersection between computers and human language -- how computers can do useful things with language. Students do not need extensive programming skills, although they are an advantage; the willingness to experiment with existing tools is, however, vital.

Rationale and Approach

This course is a survey of applications and their underlying techniques. It is intended for a wide variety of students: experience in linguistics or programming is not required. You and your classmates will have a range of skills and expertise, and the course will provide many opportunities to learn from and help each other.

Your work for this class falls into 3 categories: 1) preparation for class, 2) in-class activities and discussions, and 3) homework assignments.

Preparation: The schedule lists the topics and readings for each class meeting. Folders in the Sakai Resources/Class Materials contain materials for each class, including preparation instructions, readings, links to web materials, and discussion questions. Slides for each class will be added prior to the class meeting, in case you would like to take notes on them. Readings are taken from research literature, textbooks and collections, and manuals and handbooks. You should note any questions, comments, or ideas you have about the readings.

You are welcome to work together to prepare for class.

In-class activities: Class meetings will typically consist of 3 sections

Business – Operational questions, discussion of assignments, and other "class infrastructure".

Presentation of material -- I will highlight and review concepts and ideas from the readings, and we will discuss your questions and observations.

Activities – Activities will include "I Need to Know" presentations, demonstrations, exercises, and small-group and full class discussions.

Homework assignments: Assignments include practice with some NLP tools, "I Need to Know" presentations, a brief literature review, and a final paper or project. See the Overview of Assignments document in the Sakai Resources/Official Course Documents folder.

Keys to Success

- Plan ahead! Success in this course requires project management skills: identify milestones and deadlines, and plan your work accordingly.
- Coordinate the work schedule for this class with the schedules for your other classes, work, and other activities. You are likely to have many deadlines toward the end of the semester, so it's important for you to keep up.
- Give yourself plenty of time to prepare for each class. You will find some of the readings challenging, I know, but get as much as you can from each one. If you are not prepared for class, you will not be able to fully participate in (and benefit from) each class meeting. Especially in a small class like this, it is important for each class member to participate. It's also more interesting and more fun for us all!
- If you have any questions or concerns about the class or your work for the class, please talk to me about them. I cannot help you if I don't know there is an issue, and it is always easier to deal with problems earlier rather than later.

The most important key of all is to take advantage of this course to look at the world in a different way.

- **Enjoy and appreciate human natural language.** Pay attention to language and its context: what people say or write, the language they use, what they seek to accomplish with their language.
- **Think about the respective strengths and limitations of people and computers regarding language.** What do people do well? What do current NLP systems do well? How can we design systems that take advantage of their respective strengths, and/or compensate for their respective limitations? What new opportunities do you see for future applications of NLP?
- **Play!** There are many NLP tools and toys available online. Many research projects have websites with examples or demonstrations. Try them with different kinds of language and see what happens.

Learning Objectives

1. Students will learn about characteristics of language and its use, and understand their implications for Natural Language Processing and Natural Language Processing applications.
 2. Students will survey a broad selection of NLP applications, and be able to describe the problems or task addressed, the materials and methods used, how the applications are evaluated, and opportunities for future developments for each application.
 3. Students will learn about basic NLP tools, and use them to analyze small text corpora.
 4. Students will identify and investigate a topic of individual interest in some depth by writing a literature review, planning and completing a project or paper, and sharing the results with their classmates.
-

Course Policies

Preparation and Attendance.

I will always be prepared for class, and will start class on time. If unforeseeable circumstances prevent this for any reason, I will try to notify you beforehand if at all possible.

I expect the same of you: be prepared for class, be ready to start class on time, and let me know in advance if you must miss a class.

If you must miss a class session unexpectedly, email or call me as soon as possible (preferably within 24 hours) to explain.

Class participation contributes to your final grade; you cannot participate if you are not present. Unexcused absences or repeated tardiness will lower your grade.

You are responsible for getting notes for a missed class from a classmate.

Reading assignments, exercises, study questions, and other preparation should be done before the class for which they are assigned so you can ask questions and participate in discussions. Some of the readings are introductory or textbook-like, others are research articles, which may be more complex. Give each reading your honest effort, but don't panic if there's something you don't understand. Make a note of your questions so we can discuss them in class.

If there is something you don't understand, say so! If you don't want to ask during class, come to my office hours or contact me by email.

Electronic Devices in Class

We will be using a variety of software tools for in-class exercises and demonstrations. Many of them are web-based, and do not require downloading software, although some do. I encourage you to bring your laptop to class when convenient. If you don't happen to have a your computer with you on a day when we are using a tool, please work with another student. Those of you who do have your computer on those days, please share with those who do not.

During class, laptops, smartphones, and similar devices should be used only for class-related work. Unrelated activities such as texting, browsing, using social media, or playing games divert your attention from the class and are distracting and discourteous to others.

Please remember to mute your phone before class starts.

Assignments

Assignments must be submitted on time through the Assignment tool in Sakai.

Although you will be able to submit an assignment after the deadline, late assignments will be penalized 5% for every day it is late.

If you have a real problem submitting an assignment on time, please talk to me *before* the due date. Getting a late start on an assignment does not count as a real problem.

Pay attention to the instructions for each assignment. Be sure you have completed each part of the assignment, and proofread it before you submit it.

Start working on assignments well in advance of the due date. Do not wait until the last minute (or hour or day) to ask questions about the assignment – I may not be available for consultation.

Most of the assignments are opportunities for you to explore a topic that you are interested in and/or want to learn more about. It's never too soon to start thinking about ideas!

Policies on Academic Integrity and Diversity

Chapel Hill has had a student-administered honor system and judicial system for over 100 years. Because academic honesty and the development and nurturing of trust and trustworthiness are important to all of us as individuals, and are encouraged and promoted by the honor system, this is a most significant University tradition. More information is available at <http://studentconduct.unc.edu/honor-system>

The UNC Honor Code is in effect for all work in this course. The “Instrument of Student Judicial Governance” gives examples of actions that constitute academic dishonesty: http://studentconduct.unc.edu/sites/studentconduct.unc.edu/files/2012_2013_Instrument.pdf#academicdishonesty

Students often ask what is okay to talk about with other students and what is not. There are some specific guidelines for this course.

- I do encourage you to help each other learn the course material – your fellow students can often be a great resource for learning. For example, you may want to work together when you are reading an assigned reading for class, or discuss your thoughts in response to a question posed in the reading notes.
- You should not discuss the details of the solution to an assignment with other students, and should never copy or share answers for an assignment with other students. It is okay to talk about course material with other students, but you should not discuss solutions to pending assignments.
- All work you submit should be your own.
- You may give and receive assistance regarding the use of hardware and software. For example, you may ask or answer a question such as "how do I [fill in the blank] in NLTK?". A question such as "What is the correct number of tokens in this text paragraph?" should be addressed to me.
- Individual homework assignments are to be done **individually**. You may consult the course readings, your notes, and even other print or web sources. (Keep in mind, however, that what you find in other sources may not be consistent with what I want you to do.) You may not consult your classmates or other people; all questions should be addressed to me.
- You must sign (check) the honor statement when you submit each assignment. This confirms that you and the work you submit conform to the Honor Code.

In support of the University’s diversity goals and the mission of the School of Information and Library Science, SILS embraces diversity as an ethical and societal value. We broadly define diversity to include race, gender, national origin, ethnicity, religion, social class, age, sexual orientation and physical and learning ability. As an academic community committed to preparing our graduates to be leaders in an increasingly multicultural and global society we strive to:

- Ensure inclusive leadership, policies, and practices;
- Integrate diversity into the curriculum and research;
- Foster a mutually respectful intellectual environment in which diverse opinions are valued;
- Recruit traditionally underrepresented groups of students, faculty and staff; and
- Participate in outreach to underserved groups in the State.

The statement represents a commitment of resources to the development and maintenance of an academic environment that is open, representative, reflective and committed to the concepts of equity and fairness.

Grading Policies

Your final grade will be based on the following:

I Need to Know About	15%
Corpus Contribution	10%
Corpus Analysis	15%
Literature Review	25%
Poster Presentation	5%
Final Product	20%
Class Participation	10%

The following grade scale will be used AS A GUIDELINE (subject to any curve) for **undergraduate** students:

Grade Range	Definition*
A 90-100%	Mastery of course content at the <u>highest level of attainment</u> that can reasonably be expected of students at a given stage of development. The A grade states clearly that the students have shown such <u>outstanding promise</u> in the aspect of the discipline under study that he/she may be strongly encouraged to continue.
B 80-89.9%	<u>Strong performance</u> demonstrating a high level of attainment for a student at a given stage of development. The B grade states that the student has shown solid promise in the aspect of the discipline under study.
C 70-79.9%	A <u>totally acceptable</u> performance demonstrating an adequate level of attainment for a student at a given stage of development. The C grade states that, while not yet showing unusual promise, the student may continue to study in the discipline with reasonable hope of intellectual development.
D 60-69.9%	A <u>marginal performance</u> in the required exercises demonstrating a minimal passing level of attainment. A student has given no evidence of prospective growth in the discipline; an accumulation of D grades should be taken to mean that the student would be well advised not to continue in the academic field.
F 0-59.9%	For whatever reason, an <u>unacceptable performance</u> . The F grade indicates that the student's performance in the required exercises has revealed almost no understanding of the course content. A grade of F should warrant an advisor's questioning whether the student may suitably register for further study in the discipline before remedial work is undertaken.

*Definitions are from: <http://registrar.unc.edu/academic-services/grades/explanation-of-grading-system/> (underlining is my emphasis)

The following grade scale will be used AS A GUIDELINE (subject to any curve) for **graduate** students:

Grade Range	Definition*
H 95-99%	High Pass
P 80-94.9%	Pass
L 70-79.9%	Low Pass
F 0-69.9.9%	Fail

*Definitions are from: <http://registrar.unc.edu/academic-services/grades/explanation-of-grading-system/>

Communications

The best way to get in contact with me (other than talking to me after class or during my office hours) is by email: shaas at email dot unc dot edu. Note that I receive a large amount of email and while I try to reply to student emails within 48 hours, there are times that it may take me 2-3 days to reply. Therefore, it is important that you get started on assignments early, so there is time for me to respond to any questions you may have. I cannot guarantee that I will be able to answer last-minute questions (e.g., within 2 days of the assignment due date).

You may also call me at 919-962-8360. These are both good ways of letting me know if you can't be in class, or if you want to make an appointment with me.

If you want to discuss something we talked about in class, and assignment or some other matter, I prefer we meet in person. Please come to my office hours or make an appointment with me.

I will use the message/email facility within Sakai to send announcements to the class, as well as to individual students.

Sakai

We will use Sakai for almost all course activities. All enrolled students should have access to the UNC Sakai site for this course: <http://sakai.unc.edu/>

Course Materials

Many course materials are stored in folders in the Sakai/Resources tool.

Official Course Documents. This folder contains the syllabus, schedule, an overview of assignments, a list of useful resources, and a brief guide to reading research articles.

Class Materials. This folder contains a folder for each class meeting, labeled with the class number and date. (If a specific topic spans 2 meetings, there will be a single folder for the topic.)

Contents of each class folder may include:

- Overview: reading and other work that you should complete before the class, as well as any in-class activities you should prepare for
- Articles that aren't available on the web
- Slides that will be used during the class (usually added 1 day before the class meeting)

It is your responsibility to check each class folder enough in advance so you can prepare for the class.

Articles. This folder contains articles that are of interest to the class but aren't readily available online. If you encounter an interesting article, please let me know!

Corpus Contribution. This folder will contain the sub-corpora and documentation you will be creating in the first assignment.

Need To Know. This folder will contain the slides and recommended resources you prepare for your Need To Know assignment.

Submitting assignments

In order for you to receive credit for an assignment, it must be submitted following the submission instructions given in the assignment specifications. Most assignments will be submitted through the Sakai Assignment tool, but a couple will be uploaded to a folder in the Sakai Resources tool.

If for some reason you are unable to submit an assignment to Sakai, as a last resort you may email it to me along with a note about the problem you encountered. Then, as soon as you are

able to, it is your responsibility to submit the exact same assignment to Sakai. The email serves as a record that you tried to submit the assignment on time, but to receive credit, your assignment must be uploaded to Sakai.

Dropbox

You each have a dropbox in Sakai that is accessible only to you and me. You may store work in progress here. If you have a question about an assignment, and it would be helpful for me to see your work, you may store the draft in your dropbox, and refer to it in your emailed question. I will look at only that file to respond to your question. Do NOT submit homework by putting it into your Dropbox

Schedule
INLS 512, Applications of NLP, Spring 2014

Wednesday 2014-01-08, Class 1

Introductions, course business

What is NLP?

[Complete the Qualtrix Survey of skills and interests by Friday, 2014-01-10, 5:00 p.m.](#)

Monday 2014-01-13, Class 2

Some NLP history

Read Sparck Jones, K. (2001). [Natural language processing: a historical review](#). *Artificial Intelligence Review* 1-12, Oct 2001. <http://www.cl.cam.ac.uk/archive/ksj21/histdw4.pdf>

I also recommend these articles: speeches by two pioneers of NLP.

Sparck Jones, K. (2005) [Some points in a time](#). ACL Lifetime Achievement Award.

Computational Linguistics, 31(1), 2-14.

Kay, M. (2006) [A life of language](#). ACL Lifetime Achievement Award. *Computational Linguistics*, 31(4), 425-438.

Wednesday 2014-01-15, Class 3

A fast tour of NLP topics and issues

Read Jurafsky, D. & Martin, J. (2009). Ch. 1, Introduction. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall. 1-18.

You may also be interested in Nadkarri, P. Ohno-Machado, L., Chapman W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18, 544-551.

[Assign Need to Know](#)

[Sign up for Need to Know presentation by Friday 2014-01-17 5:00 p.m.](#)

Monday 2014-01-20, No class

Wednesday 2014-01-22, Class 4

Introducing the corpus

Read Gries, S. (2009). What is corpus linguistics? *Language and Linguistics Compass*, 3/5, 1225-1241.

Monday 2014-01-27 Class 5

Evaluation of NLP systems

Read Resnik, P. & Lin, J. (2010). Evaluation of NLP Systems. In Clark, A, Fox, C. & Lappin, S. (Eds.) *The Handbook of Computational Linguistics and Natural Language Processing*. Malden, MA: John Wiley & Sons. Ch. 11, 271-295.

Need To Know presentation

[Assign Corpus Contribution, due Monday 2014-02-10, 9:00 a.m.](#)

Wednesday 2014-01-29 Class 6

Evaluation (2); more about corpora

Read Thompson, H. (2000). Corpus Creation for Data-Intensive Linguistics. Ch. 16, 385-401. In Dale, R., Moisl, H., & Somers, H. (Eds.) *Handbook of Natural Language Processing*. New York: Marcel Dekker, Inc. Ch. 16, 385-401.

Palmer, M. & Xue, N. (2010). Linguistic Annotation. In Clark, A, Fox, C. & Lappin, S. (Eds.) *The Handbook of Computational Linguistics and Natural Language Processing*. Malden, MA: John Wiley & Sons. Ch. 10, 238-270.
Need To Know presentation

Monday 2014-02-03, Class 7

What is a word?

Words, tokens, and n-grams

Read Jurafsky, D. & Martin, J., (2009). Ch. 4 - 4.3, N-Grams. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall. 83-95.

Watch Jean-Baptiste Michel + Erez Lieberman Aiden: What we learned from 5 million books.

http://www.ted.com/talks/what_we_learned_from_5_million_books.html 14:09

Wednesday 2014-02-05, Class 8

Part Of Speech (POS) tagging

Nouns, verbs, and those pesky prepositions

Read Jurafsky, D. & Martin, J. (2009). Ch. 5-5.8, Part-of-Speech Tagging. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall. 123-163.

Need To Know presentation

Monday 2014-02-10, Class 9

Parsing

Word + Word = Chunk

Reading TBA

Need To Know presentation

Assign Corpus Analysis, due Monday 2014-02-24, 9:00 a.m.

Corpus Collection Assignment due.

Wednesday 2014-02-12, Class 10

Semantics, Negation

Read Fellbaum, C. (1998). WordNet - introduction. In WordNet: An Electronic Lexical Database. MIT Press., 1-19.

Wiegand, M. et al. (2010). [A survey on the role of negation in sentiment analysis](#). Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, 60-68.

Assign Literature Review, due Friday 2014-03-07, 5:00 p.m.

Monday 2014-02-17, Class 11

Semantics (2) -- Sentiment

Reading TBA

Wednesday 2014-02-19, Class 12

Named Entity Recognition

What's in a name?

Read Crane, G. & Jones, A. (2006). [The challenge of Virginia Banks: An evaluation of named entity analysis in a 19th-century newspaper collection](#). Proceedings of the 6th ACM/IEEE Joint Conference on Digital Libraries, 31-40.

Monday 2014-02-24, Class 13

Discourse, Rhetorical Structure Theory

Rhetorical Structure Theory [Introduction page](#); also study the [advertisement example](#).

[Bhatia's \(2004\) Move Analysis](#), Management Styles in Writing, February 2010.

Need To Know presentation
Corpus Analysis due.

Topics and readings for the rest of the semester will be determined based on class interests.

Wednesday 2014-02-26 Class 14

TBD

Need To Know presentation

Monday, 2014-03-03, Class 15

TBD

Need To Know presentation

Wednesday, 2014-03-05, Class 16

TBD

Need To Know presentation

Literature Review due 2014-03-07, 5:00 p.m.

Spring Break 2014-03-10 & 2014-03-12, no class

Monday, 2014-03-17, Class 17

TBD

Assign Final Paper or Project. Paper/Project plan due Wednesday 2014-03-26, 9:00 a.m.

Wednesday, 2014-03-19, Class 18

TBD

Need To Know presentation

Monday 2014-03-24, Class 19

TBD

Need To Know presentation

Wednesday 2014-03-26, Class 20

TBD

Need To Know presentation

Paper/Project plan due

Monday 2014-03-31, Class 21

TBD

Need To Know presentation

Wednesday 2014-04-02, Class 22

TBD

Need To Know presentation

Monday 2014-04-07, Class 23

TBD

Need To Know presentation

Wednesday 2014-04-09, Class 24

TBD

Need To Know presentation

Monday 2014-04-14, Class 25

TBD

Need To Know presentation

Wednesday 2014-04-16, Class 26

Imagining the Future of NLP: Challenges and Opportunities

Monday 2014-04-21, Class 27

Poster Session 1

Wednesday 2014-04-23, Class 28

Poster Session 2

Friday 2014-05-02, 4:00 p.m.

Final Project/Paper due