Course Structure
NLP & Question Answering
Spring 2019

Notice: This course is a graduate-level course, advanced undergraduate students can take the course in case of high motivation and having an interest in cutting-edge areas and research

---

Scheduling: TR: 11: pm - 12:15 pm
Venue: room#120 Music and Theatre Building
Office: Room #101B, and # 101C, Music and Theater Building

Assistant Professor: **Dr. Saeedeh Shekarpour**
E-mail: sshekarpour1@udayton.edu
Phone: (937) 229 3925
Office Hours: M: 3:30-4:30

Teacher Assistant: **Sunday Ngwobia**
E-mail: ngwobias1@udayton.edu
Office Hours: MW: 1 pm- 4 pm

Teacher Assistant: **Shubham Kokul**
E-mail: kokuls1@udayton.edu
Office Hours: MWF: 2 pm- 4 pm

Text Book and Slides

Dan Jurafsky and James H. Martin, "Speech and Language Processing, 2nd and 3rd Edition", Prentice Hall, 2009.

Third edition draft is available at web.stanford.edu/~jurafsky/slp3/

Course Requirements:

➔ Prior and fresh knowledge in programming especially in Java and Python.
➔ Knowledge of Machine Learning
  Since almost all NLP problems are driven by machine learning techniques, I encourage you to read machine learning textbooks:

● Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
● Tom Mitchell, "Machine Learning", McGraw Hill, 1997.

Grading Schema

Exams ……………………………………….	40%
Project ……………….……………………	60%
Assignments ……………….…………..…	10%
Class Activities …..…..………………………	extra credit
Total ……………….………………….…	100% >>

Excellent Project: In case a team comes up with a novel idea and performs a great project in the level that we can submit to a workshop/conference, thus all the team members will obtain **A**, regardless of their grades in exams and assignments

Important Statements:

- ➜ **Academic Honesty:** You are allowed to collaborate and discuss with other students or search online, but you are not allowed to copy under any circumstances, any case of plagiarism will cause F for all the involving students.
- ➜ Your class activity and presence is a **must**, around the end of the semester I will ask students with 3 and more absence to drop the course.
- ➜ Please bring your notebook to the Lab sessions.

Course Structure

Month	Week	Topic	References
January 2019	**Week 1**	Introduction Regular Expression & Text Processing	
	Week 2	Language Modeling with N-Grams	
	Week 3	Sentiment Analysis & Text Classification	
February 2019	**Week 4**	Vector Semantics part 1	
	Week 5	Vector Semantics part 2 Lab 1	
	Week 6	**Project Delivery - Part 1**	
	Week 7	Midterm Exam Part of Speech Tagging	
March 2019	**Week 8**	Grammars & Parsing	
	Week 8	**Spring Break**	
	Week 9	Information Extraction	
	Week 10	Guest Speaker	
April 2019	**Week 12**	Sentence Semantics & Semantic Role	
	Week 13	Question Answering & Chatbots Lab 2	
	Week 14	**Project Delivery - Part 2**	
	Week 15		
May 2019	**Week 16**	Miscellaneous - Q & A	

Course Project

Project Delivery. You have two deliveries for your project (i) presentation and (ii) write-up. The structure of both should follow the following structure:

1) **Title and Team Names**
2) **Introduction (Motivation and Research Question):** Which problem does your group try to solve to improve? What are the motivations behind your choice?
3) **Literature Review:** you have to make a review of the state-of-the-art papers related to your topic, what is missing from the state-of-the-art? What you can improve?
4) **Corpus Acquisition:** what are your strategies for corpus acquisition? Do you want to crawl data? Do you want to compile data? Or do you use an existing corpus?
5) **Approach:** Describe the algorithms and machine learning models used in your project.
6) **Implementation details:** which libraries did you use, how is the contribution of the team members in the implementation? Which parts of your code come from external open-source projects? Did you upload your code on GitHub?
7) **Experimental Setup:** Describe what kind of evaluation you are doing and which methods you are comparing against each other.
8) **Results:** Include a detailed comparison of different methods.
9) **Discussion and Analysis of the Results:** what are your achievements? How do you interpret the failures (error analysis)? Did you improve over the baseline? Why or why not?
10) **Future Work:** What are the potential improvements for future? What you did not have time to finish, but you think would be a useful addition to your project.
11) **The contribution of team members:** how was the team management strategies and what was challenging?

Project presentation (Phase 1): each team is given 10 minutes to present items (1), (2) and plan for the item (3).

Project presentation (Phase 2): each team is given 25 minutes to present items (1), (2) and plan for the (3).

Project write-up: all of the above items should be written in your write-up. The deadline for delivering your write-up is Tuesday 30 April, 11:59 pm. No extension is accepted under any circumstance

Project Scheduling: click on the excel sheet here, choose the proper slot which is available for your team, after you finalize your topic and project description, email me (only one member of the team) with respect to the following structure:

Email Subject: NLP: Project Scheduling

Email Body:

Team Name: ###

Project Topic: ###

Team Members: ###

Timing: ###

Project Description: one paragraph which succinctly describes your project

NLP Libraries

1. **spacy python library:** https://spacy.io/
2. **NLTK python library:** https://www.nltk.org/
3. **NLP Stanford Group:** https://nlp.stanford.edu/
4. **Apache Open NLP:** https://opennlp.apache.org/

Project Topics

The ideas for your projects can come from:

1. **SemEval 2019**
2. **Kaggle NLP Tasks**
3. **Sebastian Ruder's curated collection**
4. **Torchtext**

Classification Tasks

5. **Toxic Comment Classification**
6. **Fake News Challenge**
7. **Spam / Click-bait detection**

Information Extraction

8. **Drug-Drug Interaction Extraction**
9. **Web named entities**
10. **Twitter sequence prediction tasks**

Parsing

11. **Universal Dependencies**
12. **WikiText**

Machine Translation

13. **WMT 2018 Shared Task**

Unlabeled Data for Clustering, Language Models, etc.

14. **Wikipedia XML data**

Sentiment and Opinion Mining

15. **Stanford Sentiment Treebank**
16. **Movie reviews**
17. **Yelp Challenge**
18. **Sentiment and opinion mining datasets**

Natural Language Understanding and Inference

19. **GLUE Benchmark**
20. **RepEval 2017**
21. **The Stanford Natural Language Inference (SNLI) Corpus**
22. **Multi-Genre NLI**
23. **MedNLI**
24. **XNLI**