

GMS 6856 Introduction to Biomedical Natural Language Processing

University of Florida Department of Health Outcomes & Biomedical Informatics College of Medicine

Instructor Name: Yonghui Wu
Credit Hours: 3
Meeting Times: TBD
Semester/Year: Spring 2019
Delivery Format: On- Campus
Location: TBD
Room Number: TBD
Phone Number: (352) 294-8436
Email Address: yonghui.wu@ufl.edu
Office Hours: By appointment
Canvas URL: <http://elearning.ufl.edu>
Department Website: <https://hobi.med.ufl.edu>

PREREQUISITES: Experience of computer programming, such as python for data processing.

MAIN TEXT (required)

Manning, Christopher D., and Hinrich Schütze. Foundations of statistical natural language processing. Cambridge: MIT press, 1999.

ADDITIONAL REFERENCES

1. Cohen, Kevin Bretonnel, and Dina Demner-Fushman. Biomedical natural language processing. Vol. 11. John Benjamins Publishing Company, 2014.
 2. Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." Journal of the American Medical Informatics Association 18, no. 5 (2011): 544-551.
 3. Meystre, Stéphane, and Peter J. Haug. "Natural language processing to extract medical problems from electronic clinical documents: performance evaluation." Journal of biomedical informatics 39.6 (2006): 589-599.
 4. Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. "Natural language processing (almost) from scratch." Journal of Machine Learning Research 12, no. Aug (2011): 2493-2537.
-

DESCRIPTION OF COURSE CONTENT

This course will examine current natural language processing (NLP) methods and their applications in the biomedical domain. This introduction will cover the basic knowledge of general NLP, basic NLP tasks at different linguistic levels, NLP applications in biomedical literature and clinical text, ontologies and resources in the biomedical domain, popular NLP methods and machine learning models, commonly used NLP tools, as well as relevant computational linguistic knowledge. It will provide hands on experience with existing biomedical NLP systems. Students will gain knowledge and skills in various NLP tasks such as Named Entity Recognition, Information Extraction, and Information Retrieval.

PURPOSE OF THE COURSE

The purpose of this course is to give students an overview of natural language processing (NLP) in the biomedical domain. Students will be exposed to a variety of commonly used NLP resources and ontologies, open-source biomedical NLP frameworks and systems, as well as hands-on sessions of using existing biomedical NLP systems. Through this course, students will be able to identify biomedical NLP resources and biomedical NLP systems that useful for their future career development.

COURSE OBJECTIVES & GOALS

Upon successfully finishing this course, students will be able to:

1. Describe the differences between Biomedical NLP and general English NLP.
 2. Apply machine learning algorithms to solve biomedical/general Natural Language Processing problems.
 3. Apply the probabilistic theory and machine learning algorithms to solve biomedical/general Natural Language Processing problems.
 4. Describe the system architecture and the state-of-the-art methods for the major topics in NLP domain: Information Extraction, Text Classification, Information Retrieval.
 5. Describe the biomedical NLP systems and evaluate their performances for specific NLP tasks.
 6. Be able to use biomedical NLP systems and biomedical resources in their future study.
-

COURSE POLICIES

ATTENDANCE POLICY

Class attendance is mandatory. Excused absences follow the criteria of the UF Graduate Catalogue (e.g., illness, serious family emergency, military obligations, religious holidays), and should be communicated to the instructor prior to or on the day of absence, not later. UF rules require attendance during the first two course sessions. Missing more than three scheduled sessions will result in a failure. Regardless of attendance, students are responsible for all material presented in class and meeting the scheduled due dates for class assignments. Finally, students should read the assigned readings prior to the class meetings, and be prepared to discuss the material for each session.

For information regarding the UF Attendance Policy see the Registrar website for additional details:

<https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx>

QUIZ/EXAM POLICY

Unless otherwise stipulated, all assignments are individual assignments. Students are forbidden to collaborate or consult with one another on such assignments. Students must of course follow the University Policy on Academic Misconduct, which includes but is not limited to prohibition of plagiarism. All assignments are due at the beginning of class. NO exception.

MAKE-UP POLICY

Students are allowed to make up work only as the result of illness or other unanticipated circumstances. In the event of such emergency, documentation will be required in conformance with University policy. Work missed for any other reason will earn a grade of zero.

TERM PROJECT POLICY

The final assignment for this course is a project which includes a paper assignment, a programming assignment, and a presentation. As such, this is a collaborative project. Students are expected to work on this project in a professional manner, and are expected to clearly delineate roles and responsibilities.

UF POLICIES

UNIVERSITY POLICY ON ACCOMMODATION STUDENTS WITH DISABILITIES: Students requesting accommodation for disabilities must first register with the Dean of Students Office: <https://drc.dso.ufl.edu/>. The Dean of Students Office will provide documentation to the student who must then provide this documentation to the instructor when requesting accommodation. You must submit this documentation prior to submitting assignments or taking the quizzes or exams. Accommodations are not retroactive, therefore, students should contact the office as soon as possible in the term for which they are seeking accommodations.

UNIVERSITY POLICY ON ACADEMIC MISCONDUCT

Academic honesty and integrity are fundamental values of the University community. Students should be sure that they understand the UF Student Honor Code at <http://www.dso.ufl.edu/students.php>. You are expected and required to comply with the University's academic honesty policy (University of Florida Rules 6C1-4.017 Student Affairs: Academic Honesty Guidelines, available at <http://regulations.ufl.edu/chapter4/4017.pdf>). Cheating, plagiarism, and other forms of academic dishonesty will not be tolerated. Note that misrepresentation of the truth for academic gain (e.g., misrepresenting your personal circumstances to get special consideration) constitutes cheating under the University of Florida Academic Honesty Guidelines

NETIQUETTE – COMMUNICATION COURTESY

All members of the class are expected to follow rules of common courtesy in all email messages, threaded discussions, and chats. The first instance of clearly rude and/or inappropriate behavior will result in a warning. The second instance will result in a deduction of five percentage points (20 points) from your overall grade. The third instance will result in a drop of a letter grade (A to B, A- to B-, and so on).

Follow netiquette for online communications <http://teach.ufl.edu/wp-content/uploads/2012/08/NetiquetteGuideforOnlineCourses.pdf>

GETTING HELP

For issues with technical difficulties for E-learning, please contact the UF Help Desk at:

- Learning-support@ufl.edu
- (352) 392-HELP - select option 2
- <https://lss.at.ufl.edu/help.shtml>

Any requests for make-ups due to technical issues MUST be accompanied by the ticket number received from LSS when the problem was reported to them. The ticket number will document the time and date of the problem. You MUST e-mail your instructor within 24 hours of the technical difficulty if you wish to request a make-up. Other resources are available at <http://distance.ufl.edu/getting-help/>

COUNSELING AND WELLNESS RESOURCES

For any type of counseling needs, especially related to stress with school, work, personal life, crisis and emergency counseling, alcohol and other drug related services as well as Psychiatry wellness, you are encouraged to make an appointment with the counseling and wellness center through this link:

<https://counseling.ufl.edu/>

LIBRARY HELP DESK SUPPORT

Should you have any complaints with your experience in this course please visit <http://distance.ufl.edu/student-complaint-process/> to submit a complaint.

EVALUATION AND GRADING POLICIES

Your semester grade will be based on a combination of class involvement, homework assignments, a group project, and a final exam.

1. Class involvement: 20%
2. Homework: 25%
3. Group project: 30%
4. Final exam: 25%

Notes: Class involvement will be assessed by how students ask questions through active thinking. Your participation in discussions is an effective part of the score you receive. Students that actively participate discussions or at least ask one question (irrelevant questions not counted) per class will get the involvement score. Homework will be assigned in several ways such as applying an existing NLP package to solve a specific problem, reading a research article, or conduct literature reviews. Homework assignments, group project, and final exam will be assessed according to the quality of finishing. Instructions for each assignment, test and for the project will be carefully described when assigned.

SCALING

Letter grade	Grade points	Grade Percentage
A	4.0	95-100
A-	3.67	90-94
B+	3.33	87-89
B	3.0	83-86
B-	2.67	80-82
C+	2.33	77-79
C	2.0	73-76
C-	1.67	70-72
D+	1.33	67-69
D	1.0	63-66
D-	0.67	60-62
E	0	<60

ONLINE FACULTY COURSE EVALUATION PROCESS

Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at <https://evaluations.ufl.edu> . Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at <https://evaluations.ufl.edu/results/> Students' evaluations are important (e.g. for assessing instructors' performance on their teaching activities in relation to promotion) and they will be used to improve the course topics, materials, assignments, exams, and the instructor's teaching style.

COURSE SCHEDULE (TENTATIVE)

Week	Topic	Notes
Week 1	Introduction to Natural Language Processing: background and history, challenges, applications	
Week 2	Linguistic essentials – Word: tokenization, normalization, part-of-speech, word sense disambiguation	
Week 3	Linguistic essentials – Sentence: sentence boundary, syntactic structure, parsing, chunking	
Week 4-5	Develop corpus: annotation, tagging, agreement, popular corpora, N-gram, Language model	
Week 5-6	Named entity Recognition: Hidden Markov models, Conditional Random Fields, Deep Learning	
Week 7-8	Mining relations: temporal relation, entity relation, rule-based method, machine learning solution	
Week 9-10	Biomedical NLP resources: ICD, UMLS, RxNorm, SnoMed, Clinical trial.	
Week 11	Biomedical NLP systems: MedLEE, MetaMap, cTAKES, KnowledgeMap, CLAMP	
Week 12-13	Biomedical NLP case study: I2B2 challenge and Bio creative challenge	
Week 14-15	Group project presentations	