Course Change Request

New Course Proposal

Date Submitted: 03/13/24 2:46 pm

Viewing: CDS 321 : Elements of Natural Language Processing

Last edit: 03/14/24 10:47 am

Changes proposed by: blaisten

In Workflow

1. **CDS Chair**
2. **SC Curriculum Committee**
3. SC Assistant Dean
4. Assoc Provost-Undergraduate
5. Registrar-Courses
6. Banner

Approval Path

1. 03/13/24 3:39 pm Jason Kinser (jkinser): Approved for CDS Chair

Are you completing this form on someone else's behalf?

Yes

Requestor:

Name	Extension	Email
Estela Blaisten	31988	blaisten@gmu.edu

Effective Term: Fall 2024

Subject Code: CDS - Computational and Data Sciences

Course Number: 321

Bundled Courses:

Is this course replacing another course? No

Equivalent Courses:

Catalog Title: Elements of Natural Language Processing

Banner Title: Natural Language Processing

Will section titles vary by semester? No

Credits: 3

Schedule Type: Lecture

Hours of Lecture or Seminar per week: 3

Repeatable: May only be taken once for credit, limited to 2 attempts (N2)

Max Allowable Credits: 6

Default Grade Mode: Undergraduate Regular

Recommended Prerequisite(s):
CDS 303

Recommended Corequisite(s):

Required Prerequisite(s) / Corequisite(s) (Updates only):
CDS 101, CDS 130, CDS 230

Registrar's Office Use Only - Required Prerequisite(s)/Corequisite(s):

And/Or	(	Course/Test Code	Min Grade/Score	Academic Level	)	Concurrency?

Registration Restrictions (Updates only):

Registrar's Office Use Only - Registration Restrictions:

Field(s) of Study:

Class(es):

Level(s):

Degree(s):

School(s):

Catalog Description:
This course teaches the fundamentals of natural language processing (NLP) and natural language understanding (NLU) and helps develop necessary skills for beginner and intermediate level computational linguistics models, useful for analyzing text or speech from different human languages. This course teaches

various NLP/NLU methods, including text mining, text analyses and parsing, topic modeling, semantic similarities, vector representations of words, and gives an introduction to large language models (LLMs).

Justification:

What: This course fills in a gap of data science topics and areas of study necessary for a computational and data science degree. It teaches students the topic of NLP both theoretically (conceptually) and practically (applications).

Why: NLP has a distinct place in data science and the whole area of machine learning (ML) and deep learning (DL). While it is intricately linked to other topics in ML/DL, it is a standalone and necessary field of study in data sciences and artificial intelligence.

Does this course cover material which crosses into another department? No

Learning Outcomes:

- Students will acquire the fundamentals of NLP and NLU foundations and the current state of algorithms in the field
- Students will be able to write intermediate level code for text mining, natural language processing and speech recognition
- students will be able to create and run basic NLP/NLU models of real-world systems, such as BERT and GPT systems.

Will this course be scheduled as a cross-level cross listed section?

Attach Syllabus

CDS_321_Syllabus_March6-2024.pdf

Additional Attachments

Staffing:

Dr. Anamaria Berea, Associate professor, Department of Computational and Data Sciences

Relationship to Existing Programs:

None

Relationship to Existing Courses:

The course provides introductory coding and logical elements useful for the CDS 403 Machine Learning Applications in Science.

Additional Comments:

The course will be added to the Extended Core Courses list of the Computational and Data Sciences, BS (banner code SC-BS-CDS).

Reviewer Comments

Key: 18653

CDS 321 Elements of Natural Language Processing

1. General Information

Instructor: Dr. Anamaria Berea (aberea@gmu.edu)
Department: Department of Computational and Data Sciences
Where and when: TBD
Required Prerequisites: CDS 101, CDS 130, CDS 230
Recommended Prereq: CDS 303
Office Hours: By appointment

2. Course Description

This course teaches the fundamentals of natural language processing (NLP) and natural language understanding (NLU) by developing the necessary skills required for using beginner and intermediate level computational linguistics models and for analyzing text or speech from different human languages. This course teaches various NLP/NLU methods, including text mining, text analyses and parsing, topic modeling, semantic similarities, vector representations of words, and gives an introduction to large language models (LLMs).

3. Learning Outcomes

By the end of the course, students will:

- acquire the fundamentals of NLP and NLU foundations and the current state of algorithms in the field
- be able to write intermediate level code for text mining, natural language processing and speech recognition
- be able to create and run basic NLP/NLU models of real-world systems, such as BERT and GPT systems.

4. Recommended textbook

Natural language processing with Python: analyzing text with the natural language toolkit, by Bird, Steven, Ewan Klein, and Edward Loper, O'Reilly Media, Inc., 2009. https://tjzhifei.github.io/resources/NLTK.pdf

5. Technology Requirements

Hardware: Students need access to a Windows, Mac, or Linux laptop or desktop computer with at least 8 GB of RAM and a fast, reliable broadband Internet connection. For the programming exercises that are computationally intensive, students will be instructed how to access open source cloud systems (i.e., Google Colab) and/or the servers available in our department.

Software: This course will be using primarily Python, and/or secondarily R as a programming language. Prior intermediate level knowledge of these languages is required.

6. Course outline (Tentative)

Week	Topic	What is Due
1	Intro to NLP	
2	Language Processing and Python	Assignment 1
3	Accessing Text Corpora and Lexical Resources	Assignment 2
4	Processing Raw Text	Assignment 3
5	Writing Structured Programs	Assignment 4
6	Categorizing and Tagging Words	Assignment 5
7	Learning to Classify Text	Assignment 6
8	Extracting Information from Text	Assignment 7
9	Analyzing Sentence Structure	Assignment 8
10	Building Feature Based Grammars	Assignment 9
11	Analyzing the Meaning of Sentences	Assignment 10
12	Managing Linguistic Data and Large Language Models	
13	The Language Challenge and Future of NLU	Class Project
14	Student Presentations	
TBA	**Final Exam**	

7. Grades and their weights

Average Assignment Score	**50%**
Class Project	**30%**
Final Exam Score	**20%**

Final Mark	Corresponding Grade
>96.7	A+
93.3 – 96.6	A
90.0 – 93.2	A-
86.7 – 89.9	B+
83.3 – 86.6	B
80.0 – 83.2	B-
76.7 – 79.9	C+
73.3 – 76.6	C
70.0 – 73.2	C-
60.0 – 69.9	D
<60.0	F

8. Assignments, Class Project and Final Exam

Each assignment consists of one or two computational exercises from the topic taught during that week in class.

The final exam will be conducted in the form of a take-home, one hour asynchronous exam, to be accessed via Blackboard (or equivalent). No group work will be allowed. The honor code will be enforced in ensuring that all work turned in will be the student own work.

The class project consists of an NLP/NLU problem that students will work on throughout the semester. Students can choose the specific topic and the datasets they want to work with. The purpose of the class project is to teach students to take a NLP/NLU problem from scratch and deliver their output results in a professional-like report that enriches their portfolio of data science projects.

9. Course website

The course has a Blackboard (or equivalent(website and a GitLab repository for the programming exercises demonstrated in class. These websites are portals where students may obtain lecture notes, retrieve assignment data, review links to additional materials, and receive special announcements.

10. General University Policies:

The course abides to the Mason policies. The information faculty should add to the public syllabus is at: *https://stearnscenter.gmu.edu/knowledge-center/designing-your-syllabus* .
A set of policies follow as attachment to his document:

Mason Honor Code (quote from catalog.gmu.edu): "To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University community and with the desire for greater academic and personal achievement, we, the student members of the university community, have set forth this honor code: Student members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work."

Email policy: Mason's electronic mail provides any official information to students. Any class materials, assignments, questions, and instructor feedback should use the email Mason email. Students are responsible for maintaining their email account active, working correctly, and should check their content regularly (review details in catalog.gmu.edu).

Plagiarism policy for Internet materials: Copyright rules apply to users of the Internet who employ elements downloaded from Internet sources. Any information in the form of graphics, text, tables, or data accessed electronically and used in homework, presentations, exams, email, reports, must be cited giving credit to the pertaining sources. Even if credit is given, students must obtain permission from any copyrighted source to use any material not created by them. Inserting someone's else material in your work is stealing intellectual property. Including a link to the site URL is currently an appropriate citation.

Student privacy policy: Mason complies with FERPA by protecting the privacy of student records and judiciously evaluating requests for release of information from those records. It is not permitted for faculty to share class progress or grade information with parents/guardians under any circumstances. Student privacy policy: https://registrar.gmu.edu/students/privacy/

Academic integrity: This course embodies the value that we all have differing perspectives and ideas, and we each deserve the opportunity to share our thoughts. Therefore, we will conduct our discussions with respect for those differences. That means, we each have the freedom to express our ideas, but we should also do so keeping in mind that our colleagues deserve to hear differing thoughts in a respectful manner, i.e. we may disagree without being disagreeable. https://oai.gmu.edu/

Students with disabilities: Students with disabilities should contact the Office of Disability Services (ODS). Students requiring special accommodations should inform the instructor the first week of classes. Accommodations may be appropriate for situations that directly affect the student academic performance. ODS requires pertinent medical documentation of a physical, mental health, attention, or other health challenge. https://ds.gmu.edu/.
