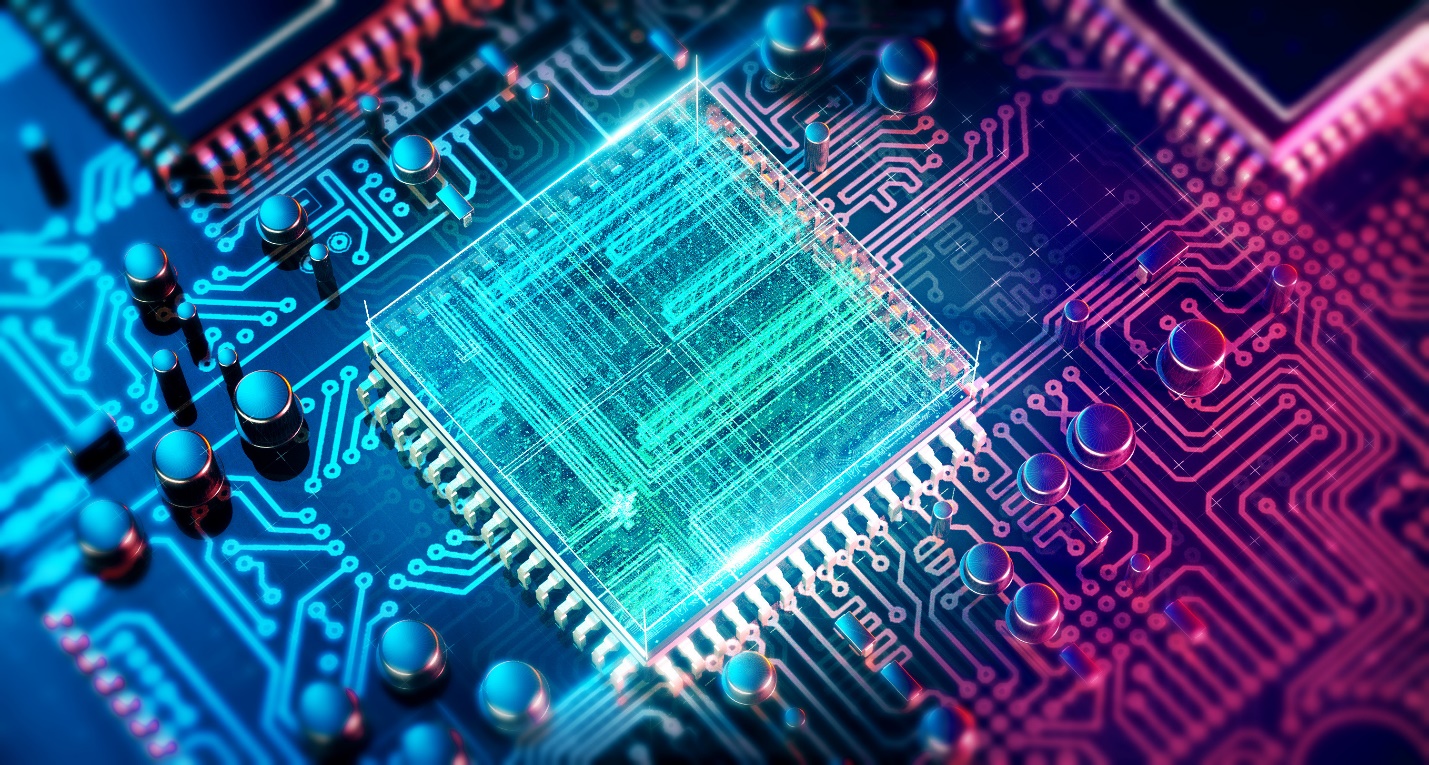
**Syllabus**

**Technical Writing and Analysis Part II: Capstone**

**Summer 2021**

**Option 2: Natural Language Processing Code Project**

****

**Course Description**

This course prepares students to succeed in the computer science field by demonstrating proper, industry standard methodologies of technical research, writing, and presentation. Students will be introduced to technical papers, white papers, and journal articles in the field of computer science and will learn how to read, decompose, and utilize them for effective integration into computer science research and project planning and development. Students will learn the proper development of white papers, technical papers, technical proposals and presentations including

how to research technical material using online databases and resources. Students will also learn proper citation methodologies including APA 6.0 and will learn how to avoid plagiarism. The final project of the class will include a technical proposal, a white paper resulting from proposal research, and a technical computer science based presentation, all with proper citations in the APA 6.0 format. Throughout the course, students will be introduced to industry standard technical and computer science terminology.

**Course Objectives**

1. Understand and apply professional writing principles in the development of computer science based journal articles and technical white papers.

2. Understand and apply sound writing principles in the development of technical proposals.

3. Understand and apply sound writing principles in the development of technical abstracts.

4. Understand and apply sound writing and visual principles in the development of technical presentations.

5. Understand how to read, decompose, and outline technical journal articles and professional technical white papers.

6. Understand and apply proper citation methodologies including the APA 6.0 format to avoid plagiarism and to add cited substantive content to technical work.

7. Understand industry related computer science and related technical terms and apply them properly in technical work products.

8. Utilize online library and technical databases to research computer science and technical literature.

9. Understand the difference between peer reviewed materials, professional materials, and non-peer reviewed materials and effectively utilize them to develop professional materials.

10. Understand and utilize industry standard resources such as ACM, IEEE, USENIX, and other professional resources.

**Course Deliverables**

1. Full set of code in Python
2. Python code second set with in code documentation
3. 5 Page professional code documentation
4. MongoDB database file
5. Brief architecture document

**Professor Dulo's Introduction**



Dr. Donna A. Dulo, the former Program Chair for the MSCS Program at Sofia University, is a mathematician, computer scientist, software/systems engineer, and legal scholar. She was with the US Department of Defense for over 32 years in both military and civilian capacities as a cyber security analyst, mathematician, computer scientist, and software engineer. Donna is a national advocate for the safety and information assurance/security of unmanned aircraft in the national airspace and speaks all over the country and works with individuals and organizations around the world on unmanned aircraft software technology, safety, and security.

Donna is a systems, software, and safety engineer for Icarus Interstellar where she designs spacecraft software systems and integrates safety, security, and resilience into spacecraft systems. Her focus is on long range spacecraft and the effects of long term stress on spacecraft systems. She is the founder of the Icarus Interstellar Center for Space Law.

At Sofia University, Donna founded and developed one of the first accredited unmanned aircraft computing master’s degree programs in the country, as well as the university’s unmanned aircraft professional certificate program. She also provides unmanned aircraft and spacecraft

software consulting services to various entities such as NASA, the FAA, as well as to technical organizations across Silicon Valley.

Donna is a legal writer and scholar in the area of cyber security and unmanned aircraft law and technology. She is the lead author and editor of the American Bar Association’s seminal text Unmanned Aircraft in the National Airspace: Critical Issues, Technology, and the Law, which was published in August of 2015. She has published extensively in the areas of unmanned aircraft law and technology, astronomy, and computer science, and has published a book on computer programming. She has been featured on several national and local media programs and articles this year including NPR, Slate, the Associated Press, and Discovery News.

**Contact Information**

Prof. Dulo can be contacted 24/7 via text message at: 831-277-2474 and via email at: Donna.Dulo@Sofia.edu.

You can speak directly to Prof. Dulo via a video Zoom session at any time, on your phone or laptop, just text her for an appointment at your convenient time and date, including weekends. She will then send you a Zoom link which will open the video chat at the time of the appointment. Feel free to schedule phone calls anytime as well.

\*\*Please do not hesitate to text at any time of day and on any day of the week (24/7)! Texting is fast, convenient, and Prof. Dulo's preferred method of contact, so that questions and issues can be resolved quickly!

**Class Schedule**

This class is held entirely online. All students are expected and required to login and participate in the class at least three times a week.

Class begins on 9 July 2021

Class ends on 20 Sept 2021

**How to Attend Class in Canvas**

Canvas is where the class action happens! Please go to the Modules portion of Canvas. The Modules are where all of the classroom materials and activities are located.

The Modules link is on the left side menu of Canvas. Each week there will be one or more modules completed. All materials presented in class will be posted in the related Module in Canvas.

To start, go to the "Module I" thread entitled “Welcome to Module I (Start Here)”. It will describe precisely what you need to do for the first Module for Week 1 of class and the learning Objectives for that Module. Each week, a new module or set of modules will be presented.

The Modules also describe the readings and topics for the week and the Learning Objectives that will be followed.

Please note that all supplemental class materials and references will be posted in “Module 0” at the top of the Modules page for easy access.

**Course Modules**

The material covered in each Module and the readings for each Module are listed under each Module heading.

The Modules in Canvas are the blueprint for the course. Each Module will consist of:

1. Welcome Thread (Start Here)

2. Readings

3. Slides

4. Videos/Supplemental Materials

5. Assignment

6. Discussion Questions

7. Module Questions Thread

**Grading Scale**

|  |  |  |  |
| --- | --- | --- | --- |
| *A* | *95 – 100%* | *C+* | *77 – 79%* |
| *A-* | *90 – 94%* | *C* | *73 – 76%* |
| *B+* | *87 – 89%* | *C-* | *70 – 72%* |
| *B* | *83 – 86%* | *D* | *60 – 69%* |
| *B-* | *80 – 82%* | *F* | *59% or <* |

**The Capstone Project**

Purpose: The purpose of the capstone project is to demonstrate a solid foundation at the graduate level of the field of computer science both in research and in application. The project allows the student to perform targeted research/coding to develop an applied solution to a real world situation or problem. The capstone also provides assessment of the student’s ability to research, write, and communicate in the area of computer science as will be required in the computing and technology industry.

Modality: This capstone is an individual project. Collaboration is not permitted with peers or classmates. All work must be the student’s own work.

Capstone Topic: Natural Language Processing

**Capstone Tasking**

Using the latest version of Python and the NLTK library, as well as other packages/libraries of yur choice (can include SPACY and SCIKIT LEARN) create a demonstration application. Create a MongoDB simple database to feed data into the application. Create documentation of the code, architecture and processes.

1. Create a simple interface that selects a random sentence (or 2 sentences) called “The Model Answer” from the MongoDB database and presents it. The interface also has an input box where a user inputs a sentence. The user had an “Enter” button.
2. The code takes both entries, runs them through a NLP pipeline to clean the data and remove stopwords.
3. The code then compares the two sentences and accounts for synonyms in the comparison. Example: car 🡪 auto, automobile, vehicle
4. The code provides a report that has:
   1. The percentage that the model answer and student answer match
   2. The number of words in each sentence (entry) before processing
   3. The original model answer and original student answer
   4. The processed model answer and student answer (after cleaning pipeline)
   5. The synonyms that could be used in the second sentence that would be a substitute for the words in the first sentence.
   6. PASS or FAIL with 70% or more being passing
5. This application simulates the major use of NLP, in this case, taking a model answer and comparing it to a student response and comparing them.
6. Utilize any freeware dictionary, as long as it does not have a cyber threat capability
7. Utilize a full NLP cleaning pipeling including:
   1. Converting text to lowercase
   2. Removing numbers
   3. Removing punctuation
   4. Removing whitespace
   5. Tokenization
   6. Stemming
   7. Lemmatization
   8. Named entity recognition
   9. Finding synonyms
   10. Spell correction
   11. Removal of stopwords (use the NLTK stopword list)
   12. Any other cleaning you feel is needed
8. Classify and parse the text as needed to properly compare the two entries
9. Use any coding techniques that you see fit.
10. Ensure that the MongoDB database properly connects to the mini application
11. Develop documentation in a 5 page report that shows the simple architecture (in any format that you like), that explains the basic processing, and explains the overall application.

**Note: This project is not a lot of code, many functions above are one liners in NLTK. The goal is complete and accurate text processing, not number of lines of code. The more concise and well structured the code is the better!**

**Recommended Submission Guidelines:**

Submit the initial code after 4 weeks

Submit architecture after 2 weeks

Submit any corrected code after 6 weeks

Submit final project at the end of class

Demo the project at the end of class (schedule with professor)

Final package due on Sept 19 (may be via email or Dropbox as canvas closes on Sept 13)

**Capstone Requirements**

1. Code works fully
2. Database integrates fully
3. Documentation is clean and professional
4. Application is accurate and professional in appearance

**Submission**

1. Full set of code in Python
2. Python code second set with in code documentation
3. 5 Page professional code documentation
4. MongoDB database file
5. Brief architecture document

**Defense**

Code demo with professor.

**Recommended Refernces**

1. Natural Language Processing by Samuel Burns. This small and inexpensive book explains it all and you really don’t need another text reference.

<https://www.amazon.com/Natural-Language-Processing-Step-Step/dp/1699028451>

1. Natural Language Processing with Python and SPACY by Vasiliev, No Starch Press

<https://www.amazon.com/Natural-Language-Processing-Python-spaCy/dp/1718500521/ref=sr_1_13?dchild=1&keywords=vasiliev&qid=1626303133&s=books&sr=1-13>

1. Python Natural Language Cookbook by Zhenya Antic

<https://www.amazon.com/Python-Natural-Language-Processing-Cookbook/dp/1838987312/ref=sr_1_1?dchild=1&keywords=zhenya+antic&qid=1626303232&s=books&sr=1-1>

1. Any NLP book that you find useful

**Academic Honesty**

Cheating and plagiarism are not tolerated in this institution (as in all academic institutions in the U.S.). If you are found to be cheating on an exam or to plagiarize material for homework assignments, you will automatically receive a grade of F for the course. If you need guidance on citing sources in order to avoid plagiarizing, please speak with the instructor.

**Learner Support**

**Library**

The Sofia University Library supports the University’s educational goals by providing a collection of library resources specifically acquired to meet the needs of our students. The library provides reference and research assistance, library research instruction, and interlibrary loan services.

The library acquires print and electronic books and periodicals in support of all academic programs. Currently the Sofia University Library has a book collection of 17,000 print volumes, 170 print periodicals, and over 30 proprietary online databases containing scholarly and curricular-based electronic books, journals articles, theses and dissertations, and electronic media. Students can access the library’s online databases and catalog via the Internet 24/7 from on- and off-campus. On-site and virtual reference service and research assistance is available. The library also offers Sofia University students the option to acquire items from other libraries through our ILL services.

Computer workstations, scanners, and printers are available in the library. Students are also able to connect to their own devices to the Sofia University wireless network. iPads and laptops (PC and Mac) are available for 24-hour checkout. In addition, the library has digital cameras, LCD projectors, and other audiovisual equipment available for short-term loan.

Library Hours

The Sofia University Library open hours are posted on the library website at https://sofia.libguides.com/homepage (Links to an external site.). Reference and research assistance is available on-site and online by appointment.

Library Contact Information

The Sofia University library staff can be reached by

Phone: (650) 388-5341

FAX: (650) 852-9780

Email: refdesk@sofia.edu

For more information and to access library online resources visit the library website at https://sofia.libguides.com/homepage (Links to an external site.)

**The Writing Lab and Dissertation Office**

The Writing Lab and the Dissertation Office are two of several academic centers at Sofia University. The main goal of these offices is to foster academic excellence across Sofia programs by supporting Sofia students, faculty, and staff in the pursuit of rigorous scholarship that furthers the Sofia mission as a dynamic and passionate learning community. The Writing Lab and Dissertation Office work together to serve the Sofia community. Email us at:

Writing Lab: lab@sofia.edu

Dissertation Office: dissertation@sofia.edu

APA Style and Citations

Students must use the Publication Manual of the American Psychological Association, 7th Edition (Links to an external site.) in preparing their written assignments. This is a new edition that will be available October 1, 2019. Please make sure you purchase the 7th Edition (Links to an external site.).

**Campus Wireless Network**

The IT Department maintains a wireless network on campus. As with any Wi-Fi network, there may be dead spots in some locations, but generally, the network is available anywhere on campus. The network is encrypted. Students should visit the Sofia University Library for assistance setting up a profile on their laptop to access the network.

**IT Help Desk**

If you need assistance logging in, accessing, or using your Sofia email, Microsoft Office Online elements, or Canvas LMS, please send an email to helpdesk@sofia.edu. Make sure to include your name and student ID in the email.

