

CIS 096L - Natural Language Processing (NLP) I

DIVISION/DEPARTMENT:	Laney - Division of Math & Science / L - Computer Information Systems	BOARD OF TRUSTEES APPROVAL DATE:	04/25/2023
ORIGINATOR:	Nguyen, Tuan	STATE APPROVAL DATE:	04/28/2023
STATE CONTROL #:	CCC000637713	CURRICULUM COMMITTEE APPROVAL DATE:	12/02/2022
		REQUISITE VALIDATION:	12/02/2022
		CURRENT EFFECTIVE DATE:	08/01/2023

1. REQUESTED CREDIT CLASSIFICATION:

COURSE TYPE: D - Credit - Degree Applicable

CB08 BASIC SKILL STATUS (PBS STATUS): N - Not Basic Skills

CB24-PROGRAM COURSE STATUS: 1 - Program Applicable

2. DEPT/COURSE NO:

CIS 096L

3. COURSE TITLE:

Natural Language Processing (NLP) I

4. COURSE:

Laney New Course

TOP NO.: 0799.00 - Other Information Technology*

5. UNITS:

Variable No

Units (Min) 3.000	Min Total
Lecture Hours (Min) 2.000	35
Lab/Studio/Activity Hours (Min) 4.000	70
Units (Max)	Max Total
Lecture Hours (Max)	0
Lab/Studio/Activity Hours (Max)	0

6. SELECTED TOPIC:

NO. OF TIMES OFFERED AS SELECTED TOPIC:

AVERAGE ENROLLMENT:

7. JUSTIFICATION FOR COURSE:

Provides an important foundational course in NLP practice through the use of available software tools and libraries; identifies the importance of data visualization in NLP; applies data visualization specific to NLP and techniques to training machine learning models. This first course in NLP will be part of the AS degree in AI and NLP certificate.

8. COURSE/CATALOG DESCRIPTION:

Fundamental concepts in Natural Language Processing (NLP): Basic understanding of NLP and its applications, NLP models and algorithms, data sets and visualization techniques, and NLP programming tools.

9. OTHER CATALOG INFORMATION:

a. Modular: No

If yes, how many modules:

b. Open entry/open exit: No

c. Grading Policy: Letter Grade Only

d. Eligible for credit by Exam: No

e. Repeatable according to state guidelines: No

f. Required for degree/certificate (specify):

g. Meets GE/Transfer requirements (specify):

h. C-ID Number:

Expiration Date:

i. Are there prerequisites/corequisites/recommended preparation for this course? Yes

10. LIST STUDENT PERFORMANCE OBJECTIVES (EXIT SKILLS):

1. Explain NLP and its applications. (I)
2. Explain structured and unstructured data in NLP data acquisition. (II)
3. Apply data curation tools and data visualization. (II)
4. Practice NLP classification. (III)
5. Use NLP models and distinguish vectorization techniques. (III)
6. Apply Various NLP data pipeline and classifiers to train machine learning models. (III, IV)
7. Create NLP applications such as Chatbot. (IV)
8. Use Python libraries in various processing methods. (III, IV, V)
9. Explain social and ethical considerations and benefits. (V)

11. COURSE CONTENT:**LECTURE CONTENT:**

- I. [15%] Natural Language Processing (NLP)
 - A. AI for Natural Language Processing
 - B. Applications of NLP
 - C. Overview of NLP data processing
 - D. Bag of words

- E. Algorithms used in NLP
- II. [20%] NLP data processing
 - A. Data acquisition for NLP
 - B. Data storage and processing
 - C. Curated data sources
 - D. NLP tools
 - E. Visualization of data
- III. [25%] Classification for NLP
 - A. Conversion of data
 - B. Term frequency-inverse document frequency (TFIDF)
 - C. Machine learning models and the sklearn library
 - D. Data pipelining
 - E. Stemming, lemmatization and tokenization
 - F. NLP model training
- IV. [20%] Language recognition application creation
 - A. Applications
 - 1. Chatbot
 - 2. Other
 - B. Applications training and basic functionality
 - 1. Neural networks
 - 2. Cosine similarity
 - 3. Natural Language Toolkit (NLTK)
- V. [20%] Algorithms and sensors
 - A. Tool set-up for data collection
 - B. Analysis of quality of sensor data
 - C. Application of sensor data
 - D. Benefits and social and ethical considerations

LAB CONTENT:

1. [15%] Describe Natural Language Processing (NLP) and applications.
2. [20%] Practice data processing relevant to data acquisition, storage, curation and visualization.
3. [25%] Implement NLP classification to allow data processing, techniques, and training.
4. [20%] Build language applications (e.g., chatbot) and training functionality.
5. [20%] Test algorithms and sensors: Setting up tools for data collection and analysis; applications stemming from sensor data .

12. METHODS OF INSTRUCTION:

- Lecture
- Lab
- Multimedia Content
- Projects
- Discussion
- Observation and Demonstration
- Visiting Lecturers
- Individualized Instruction
- Threaded Discussions
- **Other Methods:** Depending on different styles and preferences of the instructors, some of them may choose to include Observation and Demonstration, Discussions, Forums, Online Conferencing, Projects, Individualized Instruction, and more.

13. ASSIGNMENTS:

Out-of-class Assignments (List all assignments, including library assignments. Requires two (2) hours of independent work outside of class for each unit/weekly lecture hour. Outside assignments are not required for lab-only courses, although they can be given.)

Override Outside Class Hours: No

Outside-of-Class Hours (Min): 4.000

Outside-of-Class Hours (Max): 0.000

Override Outside-of-Class Hours (Min):

Override Outside-of-Class Hours (Max):

Out of class Assignment:

1. Problem sets, which are relevant to the content and levels of difficulty presented in the lectures.
2. Quizzes and/or examinations.

14. STUDENT ASSESSMENT:

- ESSAY (Includes "blue book" exams and any written assignment of sufficient length and complexity to require students to select and organize ideas, to explain and support the ideas, and to demonstrate critical thinking skills.)
- COMPUTATION SKILLS
- NON-COMPUTATIONAL PROBLEM SOLVING (Critical thinking should be demonstrated by solving unfamiliar problems via various strategies.)
- SKILL DEMONSTRATION
- MULTIPLE CHOICE
- OTHER (Describe)

- **Other:** Depending on the various teaching styles of the instructors, some of them may choose to include Online Verbal Examination, Anti-cheating Assessment (blocking online solvers, traceable identification of the questions), Multiple Choice, Matching, Calculated Multiple Choice, Multiple Answer, True/False, Drag & Drop, Multipart Question, Conditional Multipart Question, Scenarios and Simulations, Decision-making Exercises, and more.

15. TEXTS, READINGS, AND MATERIALS:

A. Textbooks:

Kochmar, Ekaterina. *Getting Started with Natural Language Processing: A Friendly Introduction Using Python*. 1 edition Pearson, 2022.

The course content and reading materials will be provided for the students in PDF format. Current up-to-date web resources will be embedded into the institutional CMS Canvas.

*Date is required: Transfer institutions require current publication date(s) within 5 years of outline addition/update.

B. Additional Resources:

Library/LRC Materials and Services:

The instructor, in consultation with a librarian, has reviewed the materials and services of the College Library/LRC in the subject areas related to the proposed new/updated course

Print Materials were reviewed? Yes

Non-Print Materials were reviewed? Yes

Online Materials were reviewed? Yes

Services were reviewed? Yes

Specific materials and/or services needed have been identified and discussed. Librarian comments:

Need recommendations from faculty for non-textbook supplementary print materials.

C. Readings listed in A and B above are:

Primarily college level

16. DESIGNATE OCCUPATIONAL CODE:

CB09 SAM Code: C - Clearly Occupational

17. LEVEL BELOW TRANSFER:

CB21 Levels Below Transfer: Y - Not applicable

18. CALIFORNIA CLASSIFICATION CODE:

CB11 California Classification Codes: Y - Credit Course

19. NON CREDIT COURSE CATEGORY:

Y - Not Applicable, Credit course

20. FUNDING AGENCY CATEGORY:

CB23 Funding Agency Category: Y - Not Applicable (funding not used to develop course)

REQUISITES AND ADVISORIES:

PREREQUISITE:

CIS 096F Introduction to Artificial Intelligence and Machine Learning

RECOMMENDED PREPARATION:

MATH 100 Mathematics for Artificial Intelligence