COURSE SYLLABUS

Course Number:	CS 174		
Title:	Special Topics in Artificial Intelligence: Natural Language Processing		
Department/Program:	DISCS	**School:**	Science and Engineering
Semester:	Second Semester	**School Year:**	2019 - 2020
Instructor/s:	Mr. Francis Bautista		
Section and Schedule:	Section B CTC214 1100 - 1230		

A. COURSE DESCRIPTION

This course provides students with the basic understanding necessary to develop and implement their own natural language processing software and algorithms in whatever field of application they may be required. This course concentrates on the fundamental techniques used to design NLP systems and the intuitions behind them.

B. COURSE LEARNING OUTCOMES

By the end of this course, students should be able to:

CLO1 - Understand the history, current state, types and limitations of NLP
CLO2 - Recall linear algebra, probability theory, graph theory
CLO3 - Understand concepts and standard processes used in transforming narratives to proper form for NLP
CLO4 - Build information retrieval algorithms that make use of basic NLP techniques
CLO5 - Develop formal models to express natural language phenomenon
CLO6 - Integrate techniques in NLP to make sense of narratives

C. COURSE OUTLINE / TIMEFRAME

CLO	WEEK / TOPIC	LEARNING OBJECTIVES	ACTIVITIES	STUDENT OUTPUT
CLO1, CLO2, CLO3	**1-2 Introduction** ● What is Natural Language Processing? ○ History of PH NLP ○ Rule-based NLP ○ Corpus Building and Annotators ● Linear Algebra Review	● Understand the history of NLP ● Identify use-cases that require NLP ● Attain a basic understanding of Linear Algebra	● Lecture ● Exercise on Linear Algebra	● Submission of exercise

CLO2	**3-4 Probability Theory and Convex Optimization** ● Probability Theory Review ● Intro to Convex Optimization and Stochastic Gradient Descent	● Explain the basics of probability theory necessary for Neural Networks ● Describe the process behind SGD and understand the intuition behind convex optimization	● Lecture ● Laboratory Exercise	● Submission of exercise
CLO3	**5-6 Word Vectors A** ● Word2Vec Definition ● Efficient Word2Vec ● Doc2Vec	● Understand the algorithms behind Word2Vec and Doc2Vec ● Implement Doc2Vec and Word2Vec using Gensim or SpaCy	● Lecture ● Laboratory Exercise	● Submission of exercise

	WEEK / TOPIC	**LEARNING OBJECTIVES**	**ACTIVITIES**	**STUDENT OUTPUT**
CLO3	**7-8 Word Vectors B** ● Global Vectors for Word Representations ● Examples ● Evaluation Metrics for Unsupervised Word Embeddings	● Explain the rationale for building word vectors ● Understand when to use specific evaluation metrics for word embeddings	● Lecture ● Laboratory Exercise	● Submission of exercise
	9 Midterms ● Project (1) specs to be announced	● Accomplish Midterm ● Assessment	● Midterm Exam ● Project submission and defense	● Exam
CLO4	**9-11 Neural Networks** ● Definition and NN Architectures ● Differential Calculus Review ● Backpropagation	● Understand the developmental history of neural networks ● Define the multiple architectures of NNs ● Refresh an understanding of differential calculus ● Understand backpropagation and its necessity	● Lecture ● Laboratory Exercise	● Submission of exercise
CLO4	**12-14 Recurrent Neural Networks and LSTMs** ● Definition ● RNN Theory ● LSTM Theory	● Explain and implement recurrent neural networks and LSTMs using TensorFlow ● Understand the theoretical architecture behind RNNs and LSTMs	● Lecture ● Laboratory Exercise	● Submission of exercise

		● Tune and modify RNN and LSTM parameters to fit specific use-cases		
CLO2, CLO4	**15 Graph Theory** ● Definition ● Graph Theory Concepts	● Understand the intuition behind representing concepts using graphs ● Understand graph traversal, clustering, searching, and ordering	● Lecture ● Laboratory Exercise	● Submission of exercise
CLO3	**16 Graphical NLP** ● Word2Vec on Graphs ● Community Detection ● Named Entity Recognition	● Understand the intuition behind representing Word2Vec on Graphs ● Implementing community detection on semantic graphs ● Apply NER algorithms using semantic graphs		
CLO6	**17 Finals** ● Project 2	● Accomplish Final Project ● Assessment	● Project submission and defense	● Project submission and defense

D. REQUIRED READING

Downloadable course materials can be accessed at the course web site, **http://moodle.ateneo.edu/ls/**
It is the responsibility of each student to regularly visit the web site for announcements, assignments, and other updates regarding the course.

E. SUGGESTED READINGS

McMillan, Michael. Data Structures and Algorithms Using C#. Cambridge University Press, 2007.
Murphy, Kevin. Machine Learning: A Probabilistic Perspective. The MIT Press, 2011.
Hastie, Trevor, et. al . An Introduction to Statistical Learning. Springer, 2009.
Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics of Machine Learning. Cambridge University Press, TBP.
Mihalcea, R., Radev, D. Graph-Based Natural Language Processing and Information Retrieval. Cambridge University Press, 2011.

F. COURSE REQUIREMENTS

Requirements	**Grade**
Exams (Midterm / Final)	30%
Projects (2)	30%
Quizzes, Seatwork, Assignments	40%
Total	**100%**

G. GRADING SYSTEM

93 – 100	A	Excellent
87 – 92	B+	Very Good
81 – 86	B	Good
75 – 80	C+	Fair
69 – 74	C	Passing
60 – 68	D	Minimally Passing
< 60	F	Fail

Notes:

- Rounding off of grades is at the discretion of the instructor. Rounding off of grades is not automatic (even if the grade is x.9999999.).
- No exemptions will be given for the final exam.

H. CLASSROOM POLICIES

1. Projects and assignments are generally submitted using the online system provided through **http://moodle.ateneo.edu/ls**/, unless otherwise specified by your instructor. All soft-copy submissions, online or otherwise, must be virus-free. Infected files will not be checked.

2. Work may be submitted late, but not later than two days beyond the specified deadline. A deduction of 5% will be applied per day late.

3. Make-up quizzes/exams will not be given. Excusable circumstances will be handled on a case-to-case basis (often involving score substitutions) but must come with a note from a doctor or from the Associate Dean for Academic Affairs (ADAA).

4. Playing games is strictly prohibited during class hours. Web browsing and reading email are also prohibited, unless done in connection with the current lecture or lab topic *and* allowed by the teacher.

5. DISCS Polices on Lab Use and Academic Integrity apply to this course. With each submission **(whether hardcopy and online)**, students must include a certification (Certificate of Authorship) that their work is substantially their own and not copied from others. In addition, students must clearly acknowledge and specify any help from outside sources such as other classmates, the Web, books, etc., that they received while doing their projects/assignments. Failure to acknowledge such may be interpreted as intellectual dishonesty. Consult the course website for details on these policies.

6. Additional policies may be implemented by the teacher, with due consultation with the students, to adapt to the class environment should any changes in semestral schedule and timing occur, and will be posted on the corresponding class website. We will never give you up, never let you down, never run around and desert you. Students are advised to be aware of such updates. A beadle will be selected for teacher-to-class communications as well. It is good practice to regularly check the class website or group, and keep in touch with the beadle.

I. CONSULTATION HOURS

Francis Bautista
TTh,1230-1330,
or by appointment

I would appreciate appointments that are set in advance.

You may contact via e-mail at **francis.bautista07@gmail.com** or at **fbautista@ateneo.edu**