UTTyler THE UNIVERSITY OF TEXAS AT TYLER SOULES COLLEGE OF BUSINESS

CSCI 5345 - Text Analytics – Spring 2026 (3 credits)

Course Description

The course explores the syntax, sentiment and hidden information contained in text documents. Students will learn how to frame text problems, choose NLP models, represent text to a computer, identify and execute machine learning techniques and evaluate model output. Includes information retrieval, natural language processing, text classification, summarization and stylometrics. Familiarity with programming is recommended. (The weekly workload may span 9 to 15 hours approximately and depends on individual students' background.)

Course Modality, Time and Location

Online Asynchronous Mode (7-week schedule); Section No.: 2026-SPRING7WK2-CSCI-5345.060.

Instructor Information

Dr. Wingyan Chung (wchung@uttyler.edu, office: COB 315.09)
Professor, Computer Science Dept.

Instructor Office Hours

Tuesday and Wednesday 9:30 am - 2:30 pm, and by appointment
Emails will normally be answered within 2 business days; emails received during Friday–Sunday will be answered by the following Tuesday. Students should NOT use the Canvas messaging system to communicate with the instructor, as Canvas lacks certain messaging features and can cause delay.

Textbook Information

The following textbooks are required throughout the semester (click the title(s) below to view the respective link(s) on a web browser).

- [S] = Dan Jurafsky, James H. Martin "Speech and Language Processing, 3rd ed. draft," (2025).
- [B] = Jens Albrecht, Sidharth Ramachandran, Christian Winkler "Blueprints for text analytics using Python," O'Reilly Media Inc., (2020). ISBN: 9781492074038.

Course Objectives

Upon successful completion of this course, students should be able to:

- Explain concepts and techniques related to natural language processing (NLP) and text analytics,
- Identify NLP algorithms and models based on the problem type and characteristics,
- Apply machine learning techniques to NLP and text analytics problems,
- Design and build NLP applications with computer programming, and
- Interpret and explain model output to business users.

Course Materials and Submissions

This class uses Canvas for dissemination of course materials, submission of assessment items, assignments, online discussions, quizzes and other class-related materials (items hosted on other sites are linked directly from Canvas). Students should check the Canvas site frequently during the semester to keep up to date about course activities (note: due dates of assessment items hosted on Canvas-linked sites may not appear in the Canvas calendar). All submissions are due by 11:59:00 pm on the due date (except otherwise stated). Students should submit their work early to avoid last-minute hassles. Email submission is NOT accepted.

Course Grading

Assessment of student performance will be based on the following:

Prefix	Category	Points
A	Assignments	64
Q	Quiz	36
Total Points		100

Grading Scale

A = 85.0 points or more
B = 70.0 to less than 85.0 points
C = 60.0 to less than 70.0 points
D = 50.0 to less than 60.0 points
F = Less than 50.0 points

Grade Composition

1. Assignment (A) – Individual, untimed, open-book, open-notes, assignments will contain objective questions, computer exercises, cases, and/or short-answer questions to help students review and practice course concepts and skills. Pause-and-resume is allowed before the submission deadline. Late submission (within 2 days after due date) will incur a 30% deduction in score. Submission is closed afterward. Missed submission will result in a zero score and cannot be made up (except for compliance with "Make-up or Extension" policy (see below)).

2. Quiz (Q) – Individual, timed, one-sitting, open-book, open-notes written quizzes will help students retain knowledge and assess learning outcomes. Objective-type questions (e.g., multiple choice, vocabularies, multiple selections, calculation, short explanations) may appear in the quizzes. No pause and no late submission is allowed. Missed submission will result in a zero score and cannot be made up (except for compliance with "Make-up or Extension" policy (see below)).

Course Policies

1. Make-up or Extension – Make-up or extension for missed submissions are available ONLY for valid reasons pre-approved by the Office of Student Accessibility and Resources (SAR) or for serious sickness / emergencies (with doctor's notices / proper documentations submitted prior to or in the same week of the deliverable) approved by the Campus Assessment, Response, & Evaluation (CARE) team. To be considered for make-up or extension (normally within 2 days after submission is closed), students must arrange with the aforementioned authorities and contact the instructor before the due date of the assessment items. *Note*: Technical issues beyond the control of the university are NOT considered as valid reasons for make-up or extension of missed submissions.

2. Student Responsibility – Students are required to take timely actions according to the course schedule, check the learning management system and their university email regularly, ensure proper technical setup in their environment (e.g., computer software and hardware, Internet access, file backup and transfer), and complete all course-related tasks on time. Students should behave properly to facilitate active class learning.

3. Academic Integrity – Any act or attempt of academic dishonesty, such as (but not limited to) plagiarism, cheating, collusion, falsifying records, and copyright infringement, is strictly prohibited and will be punished according to the university policies (e.g., §8-802).

4. UT Tyler AI Statement – UT Tyler is committed to exploring and using artificial intelligence (AI) tools as appropriate. For this course, students must complete all assessment items exclusively by themselves. When use of AI tools is permissible in specific assessment items, it will be clearly stated in the directions, and all use of AI tools must be appropriately acknowledged and cited. Otherwise, the default is that AI tools are not allowed during any stage of an assessment.

Course Schedule

Start Date	Week	Topic	Materials	Due
3/2	1	Course Introduction Words and Tokens; Python Programming Basics	Syllabus S2, B1	QS
-	-	(Spring break: 3/9–3/14 – No class)	-	-
3/16	2	N-gram Language Models Preparing Textual Data and Feature Engineering	S3 B4-5	A1
3/23	3	Document Classification and Sentiment Analysis with Logistic Regression	S4 B6-7	Q1
3/30	4	Embeddings and Vector Semantics Explaining Models and Classification Results	S5 B10	A2, Q2
4/6	5	Neural Network Language Models, RNN, LSTM Sequence Labeling for Named Entity Recognition	S6, 13, S17.3 B12	Q3
4/13	6	Transformers, Large Language Models	S7-8, S15 B11	A3, Q4
4/20	7	Masked Language Models Retrieval-Augmented Generation	S9, 11	A4 (Due: Fri.)

Remarks:
S – Course textbook chapter(s) by Jurafsky & Martin
B – Course textbook chapter(s) by Albrecht et al.
QS – Repeatable Syllabus Quiz (available in the UT Tyler Syllabus module). A satisfactory score is required to access subsequent modules.
A – Assignment
Q – Quiz