



## 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 18 hours approximately and depends on individual students' background.)

## Course Time and Room

Online Asynchronous Mode (7-week schedule)

## Instructor Information

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

## Office Hours

Tuesday and Thursday 2:30 - 4:00 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.

## Textbook Information

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

- [S] = Dan Jurafsky, James H. Martin “[Speech and Language Processing, 3rd ed. draft](#),” (2024).
- [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.

## Computer Account Access

Students will need a Patriot account and password for computer access. This information can be found at <https://www.uttyler.edu/technology-support/patriots-account/>.

## Course Materials

This class will use Canvas for course documents, slides, submission of assessment items, assignments, online discussions, quizzes and other class-related materials. Students should check the Canvas site frequently during the semester to keep up to date about course activities. 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:

Assignments (8 + 12 + 12 points)	32
Quiz (4 × 6 points)	24
Final Exam	44
<hr/> Total Points	<hr/> 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

## Course Policies

1. Assignments (A) – Individual, untimed, open-book, open-notes, assignments will contain objective questions, programming exercises, and/or short-answer questions to help students review and practice course concepts and skills. Late submission (within 2 days after due date) will incur a 30% deduction in score. Submission is closed afterward.
2. Quizzes (Q) – Individual, timed, close-book, close-notes quizzes (proctored by ProctorU) will be given periodically. Each quiz will be open for a one-week period and must be submitted by its deadline (and will be closed afterward). Missed quizzes cannot be made up without acceptable emergency-related documentation (sent to the instructor before the quiz or within 1 day after the quiz deadline). No pause is allowed in each quiz.
3. Final Exam (E) – Individual, timed, close-book / close-notes comprehensive final exam (proctored by ProctorU) is scheduled to be done within a 24-hour window. Course handouts and previous assignments will be temporarily unavailable during (and shortly before) the exam. The exam arrangement details will be provided later in the semester.
4. Make-up or Extension for Missed Submissions – Make-up or extension for missed submissions are available ONLY for valid reasons pre-approved by the [Office of Student Accessibility and Resources](#) or for serious sickness / emergencies (with doctor's notices / proper documentations submitted prior to or in the same week of the deliverable). To be considered for make-up or a short extension (normally within 2 days after submission is closed), students must email official documentations to the instructor before the due date of the assessment items.
5. Student Responsibility – Students are required to take timely actions according to the course schedule, to check the learning management system and their university email regularly, and to perform all course-related tasks. Students should behave properly to facilitate active class learning.
6. 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](#)).

7. 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

Week	Start Date	Topic	Materials	Due
1	8/26	Course Introduction Basic Text Processing; Python Programming Basics	Syllabus; S2 B1	-
2	9/2	Regular Expression and N-gram Language Models Preparing Textual Data and Feature Engineering	S2-3 B4-5	Q1
3	9/9	Document Classification with Naive Bayes Sentiment Analysis with Logistic Regression	S4, B6; Ref. [3] S5, B7	A1
4	9/16	Vector Semantics and Embeddings Explaining Models and Classification Results	S6; Ref. [1] B10	A2, Q2
5	9/23	Neural Network Language Models, RNN, LSTM Sequence Labeling for Named Entity Recognition	S7-8, S17.3 B12; Refs. [4, 5]	Q3
6	9/30	The Transformer, Large Language Models Chatbots and Dialogue Systems	S9-10, S15 B11; Refs. [2, 6]	A3, Q4
7	10/7	Masked Language Models; Word Sense Disambiguation Final Exam (opens 12 pm (10/11) – 11:59 am (10/12))	S11	E

### Remarks:

S – Course textbook chapter(s) by [Jurafsky & Martin](#)

B – Course textbook chapter(s) by [Albrecht et al.](#)

A – Assignments

Q – Quiz

E – Final Exam

### References

- [1] CHUNG, W. BizPro: Extracting and categorizing business intelligence factors from textual news articles. *International Journal of Information Management* 34, 2 (2014), 272–284.
- [2] CHUNG, W., CHEN, H., AND REID, E. Business stakeholder analyzer: An experiment of classifying stakeholders on the web. *Jrnl. of the American Society for Info. Sci. and Tech.* 60, 1 (2009), 59–74.
- [3] CHUNG, W., AND TSENG, T.-L. Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems with Applications* 39, 15 (2012), 11870–11879.
- [4] CHUNG, W., ZHANG, Y., AND PAN, J. A theory-based deep-learning approach to detecting disinformation in financial social media. *Information Systems Frontiers* 25, 2 (2022), 473–492.
- [5] LECUN, Y., BENGIO, Y., AND HINTON, G. Deep learning. *Nature* 521, 7553 (2015), 436–444.
- [6] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., ET AL. Attention is all you need. In *Advances in Neural Information Processing Systems* (2017), vol. 30, Curran Associates, Inc.