



Course	BUAN 6342
Course Title:	Natural Language Processing
Professor	Gasan Elkhodari
Term	Spring 2023
Meetings	Tuesday 7:00 PM – 9:45 PM / JSOM 2.714

### Professor Contact Information

Email Address:	<a href="mailto:Gasan.Elkhodari@utdallas.edu">Gasan.Elkhodari@utdallas.edu</a>
Mobile:	(214) 883-3670
Office Location:	JSOM 2.714
Office Hours:	By appointment only
TA Information	Joseph, Shailender

### General Course Information

*\*The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.*

Pre-requisites, Co-requisites, & other restrictions	<p>BUAN 6340 &amp; BUAN 6341: There are no exceptions</p> <p>It is expected that the learners will be able to perform the following tasks independently:</p> <ul style="list-style-type: none"><li>• Python Programming using machine learning packages e.g. Scikit-learn and data packages e.g. Pandas, Numpy</li><li>• Ability to interpret fundamental quantitative concepts in Statistics, basic Linear Algebra (for matrix operations), machine learning as well as basic English language grammar (for NLP operations)</li><li>• Perform basic management of toolkits e.g. debugging, software and package installation skills.</li></ul>
Course Description	<p>This course focuses on natural language processing and the utility of textual data to gain meaningful quantitative and actionable insights about the language (mainly English) using rule-based and statistical methods and to extract the information for real-world applications. Our eventual goal will be to create machine-learning programs that analyze and interpret human</p>

	language using classical text, social media and business text/unstructured data.
What we will cover	<p>You will learn some very practical, state-of-the-art, relevant applications in a hands- on manner:</p> <ul style="list-style-type: none"><li>• Unstructured Data Sourcing and Complication</li><li>• Text Pre-processing in a realistic setting</li><li>• Regular Expressions</li><li>• Text Visualization</li><li>• Topic Modeling</li><li>• Text Summarization</li><li>• Text Similarity</li><li>• Content Search</li><li>• Sentiment Analysis</li><li>• Information Extraction and Entity Relationships</li><li>• Text Based Classification and Clustering</li></ul> <p>The course uses Python’s NLTK (Natural Language Toolkit) and spaCy as the primary vehicle. While this is an applied course, we will also focus on the key concepts behind linguistics, text processing and the overall theory of NLP so that the learners will appreciate the field of NLP. Topics include:</p> <ul style="list-style-type: none"><li>• Various elements of linguistic understanding e.g. lexical, syntactic, and semantic</li><li>• Understanding the statistical/dynamic properties of language such as n- grams, minimum edit distance, TFIDF etc.</li><li>• Rule-based and data-driven approaches to building language models.</li><li>• Machine learning for understanding language.</li><li>• Application of NLP to real-world problems</li></ul> <p>The theoretical understanding will be supplemented with hands-on assignments and practice using Python Jupyter notebooks with additional Python packages such as NLTK, scikit-learn, pandas, numpy, matplotlib among others.</p> <p>PyTorch:</p> <p>PyTorch is a powerful open-source machine learning library that provides a wide range of functionality for building and training neural networks.</p>

	<p>KNIME Analytical tool:</p> <p>KNIME (Konstanz Information Miner) is an open-source data analytics platform that allows users to visually create and execute workflows for data processing, analysis, and modeling. It is designed to be user-friendly and easy to use, even for users with little to no programming experience.</p>
<p>Learning Outcomes</p>	<p>By the end of this course, students will be able to:</p> <ol style="list-style-type: none"> <li>1. Understand the use of regular expressions to detect patterns in text.</li> <li>2. Understand text preprocessing using general techniques such as tokenization, stemming, lemmatization, stop-word removal, spell correction and word- frequency analysis.</li> <li>3. Understand NLP specific techniques such as parsing, entity extraction, POS tagging and NER.</li> <li>4. Apply machine learning, NLP and text processing techniques to build analytical applications to process unstructured data-sets.</li> <li>5. Use multiple NLP packages e.g. NLTK, VADER, TextBlob, Gensim</li> <li>6. Build your own Corpora for NLP applications</li> <li>7. Apply an array of NLP operations to large and realistic datasets</li> </ol>
<p>Required Texts &amp; Materials</p>	<p>For this course, we will be using the below required textbook.</p> <ol style="list-style-type: none"> <li>1. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, by Bird, Klein and Loper (2nd edition). This book is freely available online. This book is available free. We will be referring to the Python 3.X version.</li> <li>2. Srinivasa-Desikan, Bhargav, <i>Natural Language Processing and Computational Linguistics</i>, Pakt 2018</li> </ol>

Week #	Week Start Date	Topic	Description	Lab #
1	Jan 17th, 2023	Class Introduction  Introduction to NLP and NLP Libraries	Overview of NLP and its applications; Basic concepts and techniques in NLP; Understanding the different NLP tasks and their challenges; Introduction to spaCy and NLTK, two popular libraries for natural language processing in python	1
2	Jan 24th, 2023	Text Preprocessing	Tokenization and stemming; Stopword removal and lemmatization; Text cleaning and normalization; Using KNIME for data preprocessing tasks	2
3	Jan 31st, 2023	Introduction to PyTorch for NLP	Basic concepts in PyTorch, including tensors and gradients; Getting started with PyTorch for NLP; Exploratory Data Analysis (EDA) of text data using PyTorch	3
4	Feb 7th, 2023	Feature Engineering	One-hot encoding; Count vectorization; Term frequency-inverse document frequency (TF-IDF). Using KNIME for data preparation, feature engineering, and visualization	4
5	Feb 14th, 2023	Sentiment Analysis	Understanding the problem; Implementing a simple sentiment analysis model; Evaluating the model and interpreting the results; Using KNIME for model evaluation and interpretation	5
6	Feb 21st, 2023	Word Embeddings	Understanding word embeddings; Training word embeddings; Using pre-trained word embeddings; Using KNIME for embeddings visualization	6
7	Feb 28th, 2023	Named Entities Recognition (NER)	Understanding the problem; Implementing a simple NER model; Evaluating the model and interpreting the results; Using KNIME for NER evaluation and interpretation	7
8	Mar 7th, 2023	Midterm Exam at the Testing Center		
9	Mar 14th, 2023	Spring Break		
10	Mar 21st, 2023	Text Classification	Understanding the problem; Implementing a simple text classification model using PyTorch; Evaluating the model and	10

Week #	Week Start Date	Topic	Description	Lab #
			interpreting the results; Using KNIME for text classification evaluation and interpretation	
11	Mar 28th, 2023	Advanced NLP Models	Introduction to sequence-to-sequence models and attention mechanisms; Implementing a simple machine translation model; Understanding the transformer architecture; Using KNIME for advanced NLP models evaluation and interpretation	11
12	Apr 4th, 2023	Advanced NLP Applications	Dialogue generation using seq2seq models; Text summarization using extractive and abstractive methods; Using KNIME for advanced NLP applications evaluation and interpretation	12
13	Apr 11th, 2023	NLP in Practice	Best practices in NLP; Deploying NLP models in production; Troubleshooting common issues; Using KNIME for deployment, monitoring and maintenance	13
14	Apr 18th, 2023	Deep Learning for Text	Understanding deep learning architectures for text, such as RNNs, LSTMs, and GRUs; Implementing these architectures using PyTorch for tasks such as text generation	14
15	Apr 25th, 2025	Keras and spaCy for Deep Learning	Introduction to Keras, a high-level neural networks API in Python; Using Keras and spaCy for advanced deep learning tasks such as text classification and named entity recognition	15
16	May 2nd, 2023	Project Presentation		
17	May 9th, 2023	Final Exam at the Testing Center		

### Course Policies

- The Homework-Labs are to be completed in class and are due by the end of class, unless otherwise stated in eLearning.
- Makeup Exam: There is no makeup exams. In case of medical emergency, a medical report is required including physician information.
- Missing exam: Any missing exam without medical report will be graded as Zero.

- Assignments must be submitted through eLearning. Emailed submissions are not accepted.
- Late Assignments: Subject to 10% penalty, 20% penalty after the third day.
- Class Attendance: Students who fail to attend class regularly are inviting scholastic difficulty. Absences may lower a student's grade where class attendance and class participation are deemed essential by the instructor.
- UTD Syllabus Policies and Procedures: Please visit <https://go.utdallas.edu/syllabus-policies>
- Cheating will not be tolerated. When I find evidence of cheating, the documentation is turned over to the Office of Community Standards and Conduct. (<https://www.utdallas.edu/conduct/dishonesty/>)

### **Grading scale**

<b>Top 25% Students</b>	<b>A</b>
<b>Next 25% Students</b>	<b>A-</b>
<b>Next 25% Students</b>	<b>B+</b>
<b>Remaining 25% Students</b>	<b>B and below</b>

### **Calculated Grade Weights\*\***

- Labs / Assignments (15%)
- Exam 1 – (30%)
- Exam 2 – (30%)
- Group Project (10%)
- Class attendance (10%)
- Class participations and engagements (5% - Hidden)

*\*\*The calculated grade weights are subject to change at the discretion of the Professor.*

### **Classroom citizenship**

- Cell phone use is not allowed during class or exam.
- eLearning will be used for class content.
- Slides and other class materials will be posted after class is held.
- Class announcements (e.g., change in assignment dates) will be posted in the eLearning announcements. It is the students' responsibility to regularly check the announcements (typically by having the announcement automatically forwarded to their email accounts).

