# Course Introduction

## CS4742 Natural Language Processing
## Lecture 00

Jiho Noh

Department of Computer Science
Kennesaw State University

CS4742 Summer 2025 [1]


[1] This lecture is based on the slides from Dr. Hafiz Khan at KSU.

1 Introduction to NLP

2 NLP History

3 NLP Applications

# Course Description

**This course introduces the theory and practice of Natural Language Processing (NLP).** It covers basic text processing techniques such as *regular expression*, *similarity matching*, and *semantic vector representation of text*. This course also covers modern NLP techniques for computers to understand natural language using **machine and deep learning techniques**.
The modern NLP techniques include designing and developing NLP systems and applications such as *language models*, sequence labeling, *machine translation*, *question answering*, and *summarization*. The course covers understanding, designing, and developing basic to advanced NLP techniques to solve natural language-related problems/tasks.

# Course Learning Outcomes

1. Demonstrate an understanding of basic concepts in the field of natural language processing (NLP).
2. Understand the various ML/DL-based NLP techniques for designing and implementing key components of related systems.
3. Demonstrate knowledge of Advanced NLP application systems.
4. Understand the implementation of different advanced NLP-related techniques (e.g., explaining, developing codebase with Python).
5. Research and critique computing literature, formulate/identify problems, develop a research project, and write a technical report/paper

# Course Topics and Structure

- Syllabus on D2L

# What is Natural Language Processing?

- Natural Language Processing (NLP) is an interdisciplinary field that studies how to process, analyze, or generate natural language text
  - **Subject:** natural language
  - **Research Tools:** computer science
  - **Modeling Tools:** mathematics, statistics, linguistics
- Realted areas:
  - Computational Linguistics (CL)
  - Language Technology
  - Natural Language Understanding (NLU)
  - Human Language Technology (HLT)

# Why Study NLP?

- Centrality of Natural Language
  - A primary (and natural) mode of human communication
  - Representation for most recorded human knowledge
  - A very rich and flexible representation (when compared to most formal representations)
- Language barriers
  - Human-human: Machine Translation
  - Human-computer: NL human machine interfaces, e.g., question answering and chatbots

# Knowledge Requirements for NLP

- **Phonetics and Phonology**: Sounds of language, for speech recognition per se.
- **Morphology**: Structure of words, for spelling correction, etc.
- **Syntax**: Structure of sentences for parsing.
- **Semantics**: Meaning of words and sentences.
- **Pragmatics**: Meaning of utterances in context.
- **Discourse**: Meaning of larger units of text, larger than a single utterance.

# Why is NLP hard?

**Ambiguity**:

- **Lexical ambiguity**: A word can have multiple meanings (e.g., bank)
- **Syntactic ambiguity**: A sentence can be parsed in multiple ways
- "I made her duck"
  - I cooked a duck for her.
  - I *made* (created) a duck for her.
  - I cooked a duck she owned.
  - I caused her to lower her head quickly.
  - I turned her into a duck.

1 Introduction to NLP

2 NLP History

3 NLP Applications

# 1950–1970

**Mid 1950's – Mid 1960's**

- Mostly rule-based systems and linguistic approaches.
- Research focused on machine translation, and they thought it would be easy.

**Mid 1960's – Mid 1970's: A Dark Age**

- People realized that machine translation is hard, and most of the research was not successful.
- The field was in a dark age, with little progress.

# 1970–2000

## 1970's and early 1980's

- *Revival of NLP* with a focus on *knowledge-based systems*.
- The emphasis was still on *rule-based* systems, *expert systems*, and *symbolic AI*.

## 1970's and early 1980's

- Statistical NLP emerged as a new approach. **Data-driven** methods started to win over rule-based methods.
- "Whenever I fire a linguist, our machine translation performance improves." – *Fred Jelinek, IBM Research*

## 2000's

- More data, more computing power, richer linguistic representation starts to find its way into NLP.

# 2010's – Recent Years

**2010's**

- Emergence of *deep learning* and *neural networks* revolutionized NLP.
- Vector representations of words (e.g., Word2Vec, GloVe) became popular.
- Large language models (LLMs) like BERT, GPT, and their successors were developed.
- New architectures like Transformers became the standard for NLP tasks.

1 Introduction to NLP

2 NLP History

3 **NLP Applications**

# NLP Applications

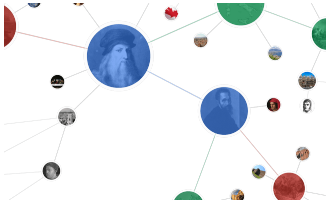
- Search engines
- Language translation
- Text summarization
- Classifying text (e.g., Sentiment analysis)
- Question Answering
- Spoken Dialog Systems (e.g., Chatbots)
- Social media analysis
- Information extraction, etc.

# Example NLP Tasks

- Text Processing and Parsing (Module2)
- Name Entity Recognition (Module 4)
- Parts-of-Speech Recognition (Module 4)
- Machine Translation (Module 5)
- Question Answering (Module 6)
- Keyword Extraction, Document Summarization (Module 7)
- ...

# Knowledge & Information Extraction

- Knowledge/Information extraction: Extracting structured information from unstructured text.
- Knowledge Graphs (KGs) organize information in a graph structure, connecting entities and their relationships, which are extremely useful for many NLP applications.



The Google Knowledge Graph is a knowledge base that enables Google to provide more relevant search results by understanding the relationships between entities and immediate factual answers to your questions.

# Machine Translation

- Machine Translation (MT): Automatically translating text from one language to another.
- **Rule-based MT**: Early systems relied on linguistic rules and dictionaries.
- **Statistical MT**: Used statistical models to learn translation patterns from bilingual corpora.
- **Neural MT**: Modern systems use deep learning models, such as sequence-to-sequence models and Transformers, to achieve high-quality translations.

*Google Translate*

- Over 500 billion words translated daily
- Supports over 100 languages, covers 95% of the world's population
- Uses neural machine translation (NMT) for most languages