



UiT The Arctic University of Norway

## Natural Language Processing

### Week 5: Natural Language Processing

Shayan Dadman - Phd Candidate, UiT Narvik  
Room: D3430  
Email: [Shayan.dadman@uit.no](mailto:Shayan.dadman@uit.no)

Andreas Dyrøy Jansson - Phd Candidate,  
UiT Narvik  
Room: C3190  
Email: [Andreas.d.jansson@uit.no](mailto:Andreas.d.jansson@uit.no)

*Department of Computer Science and Computational Engineering*



# What is natural language?

- It is the way humans communicate with each other using text and speech.
- Indeed, we are surrounded by text:
  - Signs
  - Messages
  - Emails
  - Manuals
  - Web pages
  - Etc.



# Natural Language Processing

- It is a branch of computer science, artificial intelligence (AI)
- It is concerned with machine ability to process and understand text and spoken words
- It combines computational linguistics and statistical modeling.



# Natural Language Processing

- Linguistics
  - It is a scientific study of language
  - Includes grammar, semantics, and phonetics
- Computational linguistics
  - It is a modern study of language using computer science
  - Big data and fast computers
- Statistical natural language processing
  - Similar to computational linguistics, but more engineer-based
  - It more focused on developing tools for machine translation, question answering, etc.



# What are the challenges?

- Natural language processing is still an unsolved task.
- It is hard because:
  - Text has no specific structure:
    - words with different meaning according to the context.
      - I *ran* to the store because I *ran* out of milk
    - Synonyms
    - Irony and sarcasm
    - Domain-specific language
  - It is messy
    - Errors in text and speech



# NLP Tasks

- Speech recognition:
  - known as speech-to-text, is the task of converting voice data into text data, e.g., voice commands.
- Sentiment analysis:
  - To extract subjective qualities such as attitudes, emotions, sarcasm, etc.
- Natural language generation:
  - To produce natural language. For example, weather report, chatbots, image captions, etc.



# NLP Tools

- Natural Language Toolkit (NLTK)
  - It is an open-source python library to attack NLP tasks.
  - It contains libraries such as, text preprocessing, tokenization, parsing, classification, etc.
  - It also includes sample datasets
- Statistical NLP
  - It uses the machine learning and deep learning algorithms
  - Using these methods, the NLP systems can learn from huge amount of data.
  - Methods such as:
    - Convolutional neural networks (CNN)
    - Recurrent neural networks (RNN)



# NLP Applications

- Here are some use cases of NLP in real world:
  - Spam detection:
    - Initially, this is a text classification task to scan emails for related language. For instance, overuse of financial terms, bad grammar, threatening language, etc.
  - Machine translation:
    - Best example is Google Translate. It is not about the replacing the words but to capture accurately the meaning and tone of the input language and maintain same characteristics in the output language.
  - Chatbots:
    - It is a combination of speech recognition and natural language generation. Examples are Apple's Siri and Amazon's Alexa.





# NLP Applications

- Social media sentiment analysis:
  - Here, sentiment analysis can be used to analyze language of posts, responses, reviews. Furthermore, the companies can use this information for product design, advertisements, etc.
- Text summarization:
  - To summarise the huge volume of text to provide useful context for research databases, busy readers, etc.

