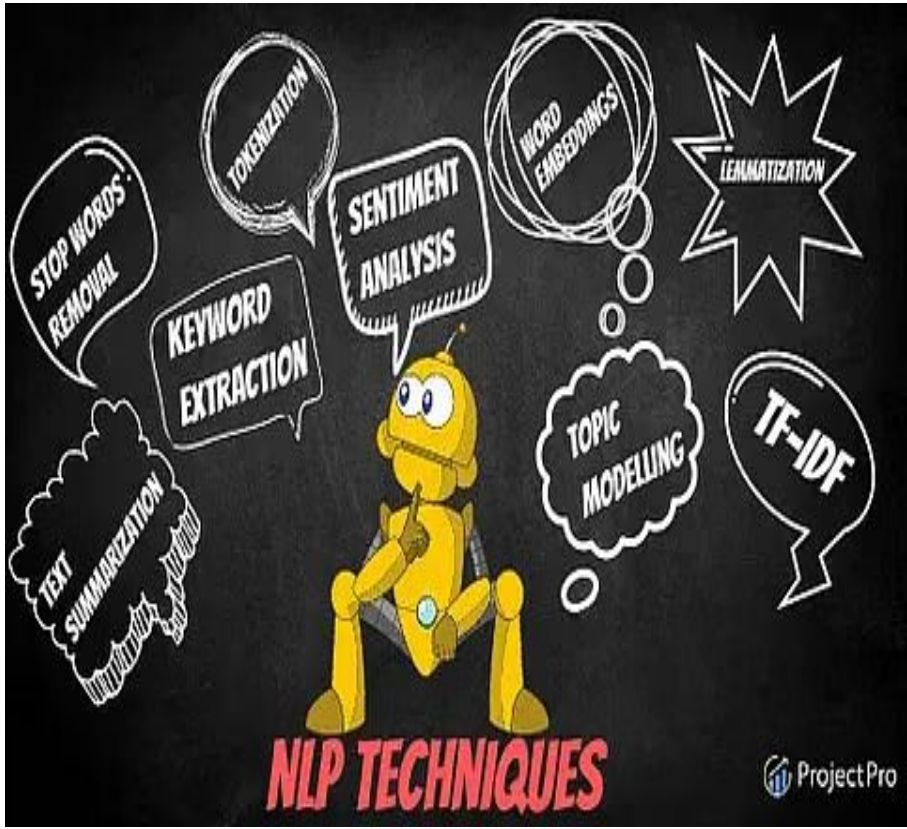
Building Large Language Model Applications

Natural Language Processing Techniques



Hamza Farooq
Dr. Saima Hassan

Learning outcomes

- Text Preprocessing
- Common NLP Tasks
- NLP Ambiguities
- Conclusion

Recap: What is NLP?

NLP helps computers **understand**, **process**, and **interact** with human language.

Some key applications of NLP are:

Chatbots & Virtual Assistants – AI-powered assistants like Siri, Alexa, and Google Assistant.
Machine Translation – Tools like Google Translate convert text between languages.
Speech Recognition – Converts spoken language into text (e.g., voice typing, transcription services).
Sentiment Analysis – Determines if text expresses positive, negative, or neutral sentiment (e.g., product reviews, social media analysis).
Text Summarization – Generates concise summaries of long documents or articles.
Spam Detection – Filters out spam emails using NLP algorithms.
Grammar & Spell Checking – Tools like Grammarly improve writing by detecting errors

Text Preprocessing in NLP

Text Preprocessing in NLP

Why Text Data Matters:

- Found in sources like customer reviews, social media, and research papers
- Helps solve real-world business problems

The Need for Text Preprocessing:

- Raw text data is often messy and unstructured
- Preprocessing cleans and organizes text for better analysis and predictions

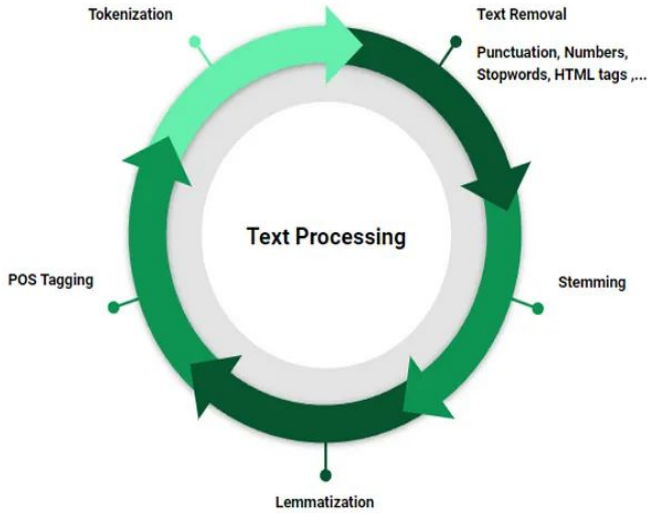
Text Preprocessing in NLP

Text Preprocessing

- Essential for cleaning and structuring text
- First step in building NLP models
- Helps improve model accuracy and efficiency

Common Preprocessing Steps:

1. Remove Punctuation
2. Remove URLs
3. Remove Stop Words
4. Lowercasing
5. Tokenization
6. Stemming:
7. Lemmatization



https://ayselaydin.medium.com/1-text-preprocessing-techniques-for-nlp-37544483c007

Text Preprocessing in NLP

Common Preprocessing Steps:

Remove Punctuation

Punctuation marks (like periods, commas, exclamation points, and emojis) often don't add value to text analysis.

Example:

- Input: "Hello, how are you doing today?"
- Output: "Hello how are you doing today"

Text Preprocessing in NLP

Common Preprocessing Steps:

Remove URLs

URLs usually don't provide useful information for NLP models.

Example:

- Input: "Visit our website at https://example.com for more info."
- Output: "Visit our website for more info."

Text Preprocessing in NLP

Common Preprocessing Steps:

Remove Stop Words

Definition: Stop words are common words (e.g., “the,” “is,” “and”) that are often removed during NLP tasks as they don’t carry significant meaning.

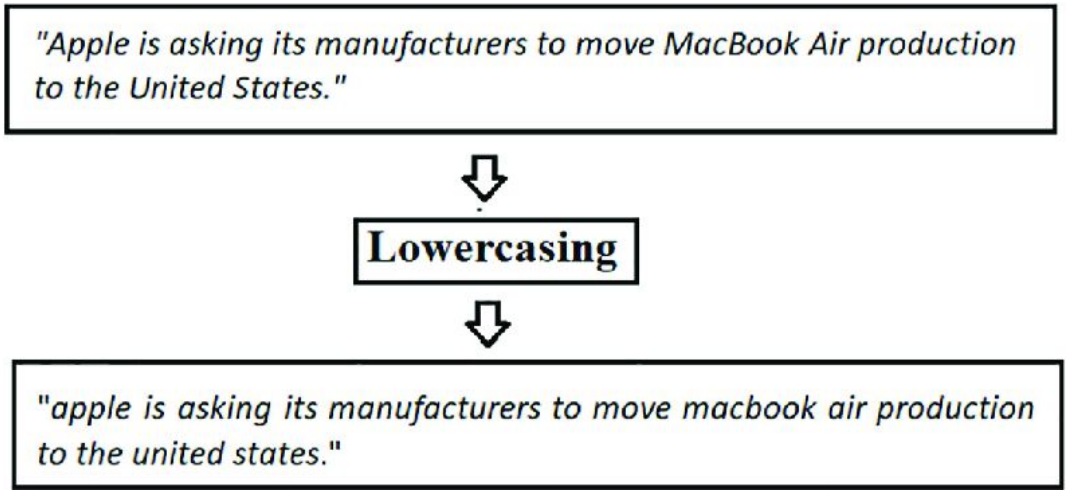
```
AI powers many modern technologies from virtual assistants to autonomous vehicles.
After Stopword
['AI', 'powers', 'many', 'modern', 'technologies', 'virtual', 'assistants', 'autonomous', 'vehicles', '.']
```

Text Preprocessing in NLP

Common Preprocessing Steps:

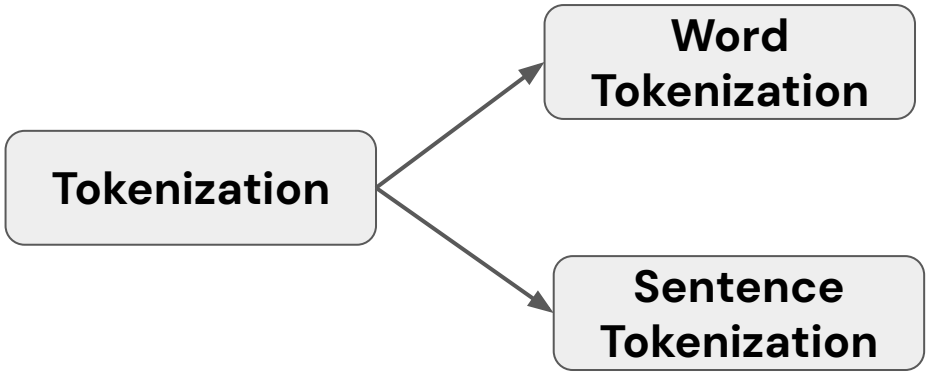
Lowercasing

In this step, the text are converted to lowercase. This step is implemented so that the algorithm does not treat the same words differently in different situations.



Text Preprocessing in NLP

Common Preprocessing Steps:



Definition: Tokenization is the process of breaking complex data like paragraphs into simple units called tokens.

```
The quick brown fox jumped over the lazy dog!
Word tokenizer
['The', 'quick', 'brown', 'fox', 'jumped', 'over', 'the', 'lazy', 'dog', '!']

I was once happy. Now that sense of joy seems to have faded.
Sentence tokenizer
['I was once happy.', 'Now that sense of joy seems to have faded.']
```

Text Preprocessing in NLP

Common Preprocessing Steps:

Stemming

Definition: process of reducing a word to its root or base form by removing suffixes or prefixes, often using a simple heuristic approach. The resulting root word may not always be a valid dictionary word.

Example:

Words: "running", "ran", “drove”, “better”

Stemming: **run**, **ran, drove, better**

Text Preprocessing in NLP

Common Preprocessing Steps:

Lemmatization

Definition: Lemmatization reduces a word to its dictionary or base form (lemma) while ensuring that the result is a valid word. It uses a vocabulary and considers the word's context (e.g., part of speech).

Example:

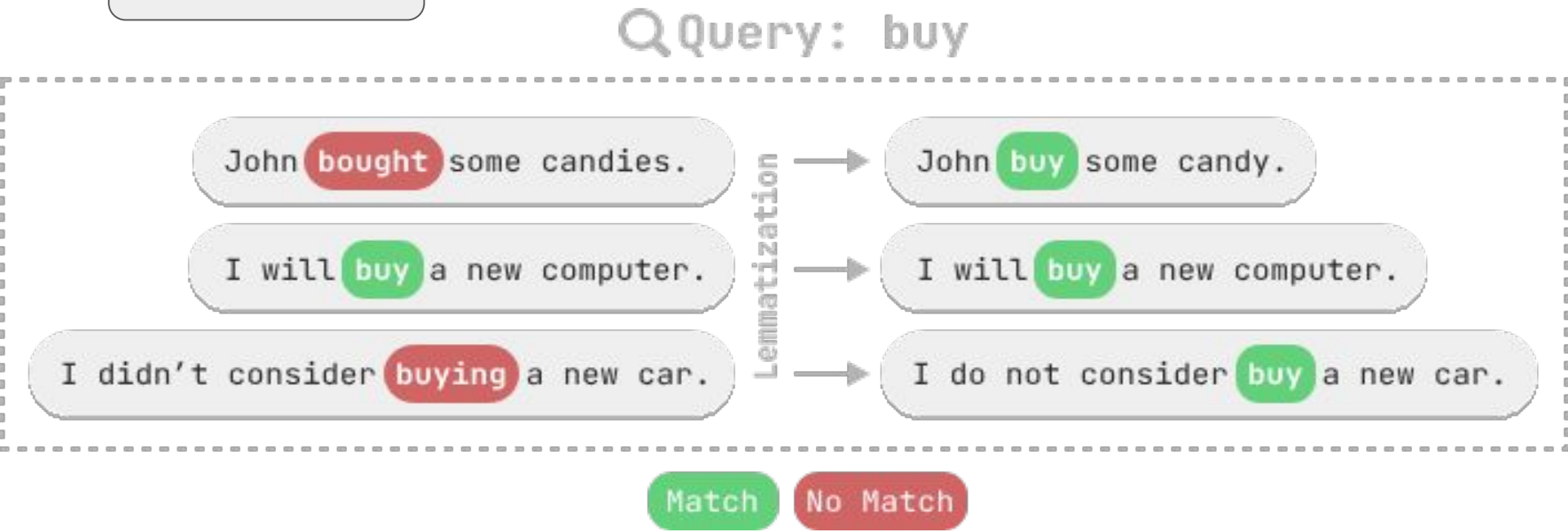
Words: "running", "ran", “drove”, “better”

lemma: run, run, drive, good

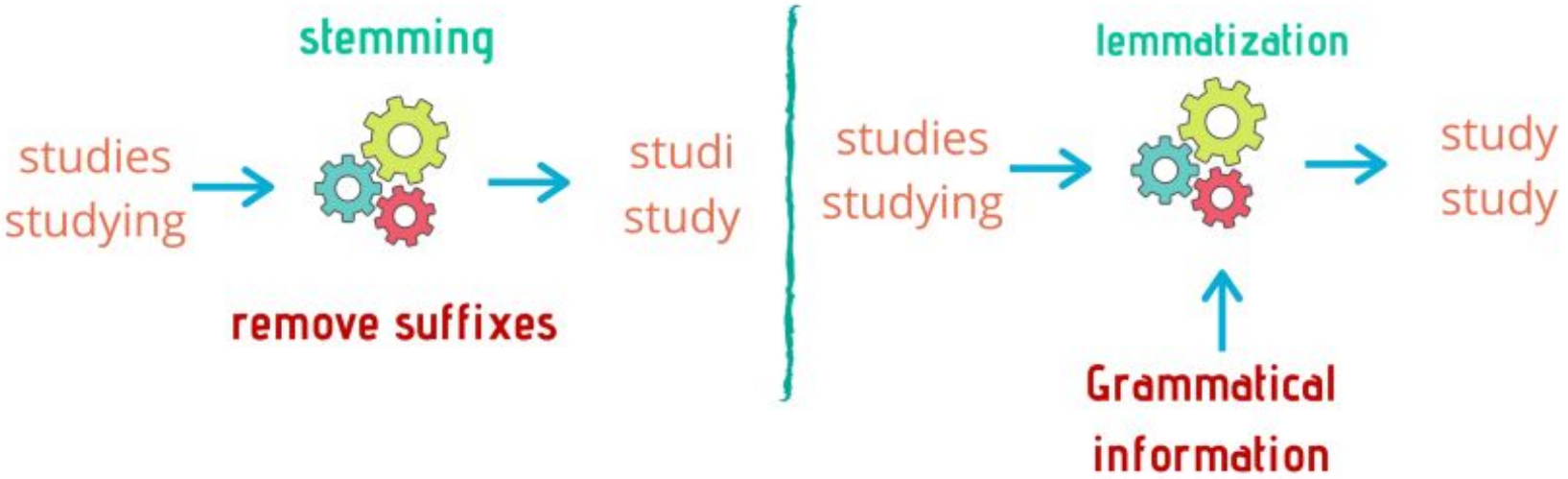
Text Preprocessing in NLP

Common Preprocessing Steps:

Lemmatization



STEMMING VS. LEMMATIZATION



Benefit: Reducing Complexity of Text Data

Lemmatization simplifies text data by reducing words to their lemmas, which decreases the vocabulary size and enhances computational efficiency in NLP tasks. This simplification makes algorithms faster and more scalable, aiding text classification by reducing unique tokens and enhancing sentiment analysis by consolidating forms of sentiment-bearing words.

Benefit: Improving the Model's Performance

The use of lemmatization leads to more uniform datasets, which can significantly improve the learning process and performance of NLP models, especially in deep learning. This benefits models like BERT or GPT by enabling better generalization over text data and supports feature extraction by effectively identifying relevant topics or keywords.

Are there any drawbacks?

One major challenge of lemmatization is the need for extensive dictionaries and complex morphological analyzers, which can make lemmatization computationally expensive compared to simpler methods like stemming.

Common NLP tasks

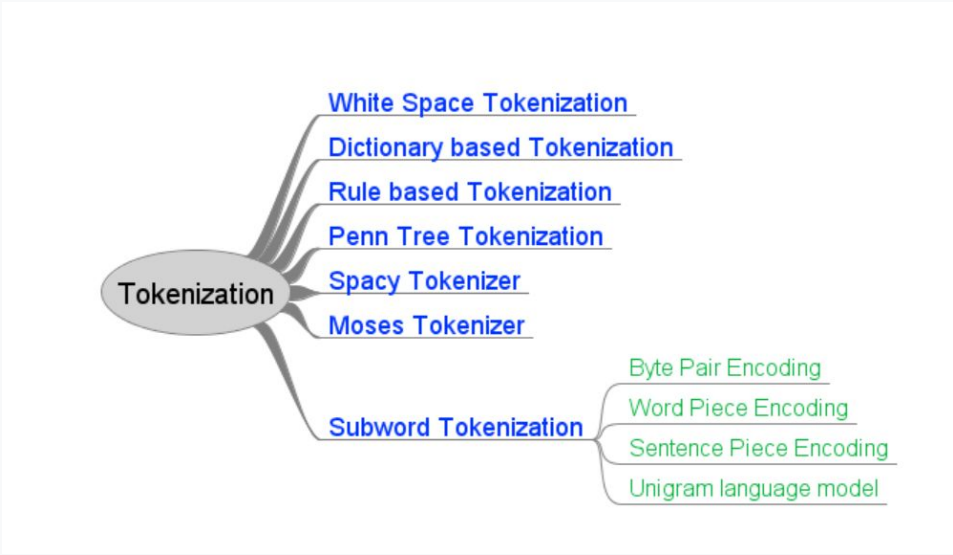
Common NLP tasks

- **Tokenization**
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- **Tokenization** is the process of breaking down a text into individual units called tokens.
- **Tokens** are typically **words**, but can also be phrases or even individual characters, depending on the application.
- Tokenization is a crucial step in natural language processing tasks such as **machine translation, sentiment analysis, and named entity recognition.**

Common NLP tasks

- **Tokenization**
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



Common NLP tasks

- **Tokenization**
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Here are some examples of tokenization in NLP:

- **Word Tokenization:**
  - Input: *"Natural Language Processing is amazing!"*
  - Output: ["Natural", "Language", "Processing", "is", "amazing", "!"]
- **Sentence Tokenization:**
  - Input: *"AI is transforming the world. NLP plays a big role."*
  - Output: ["AI is transforming the world.", "NLP plays a big role."]
- **Subword Tokenization:**
  - Input: *"unhappiness"*
  - Output: ["un", "happiness"]
- **Character Tokenization:**
  - Input: *"AI"*
  - Output: ["A", "I"]

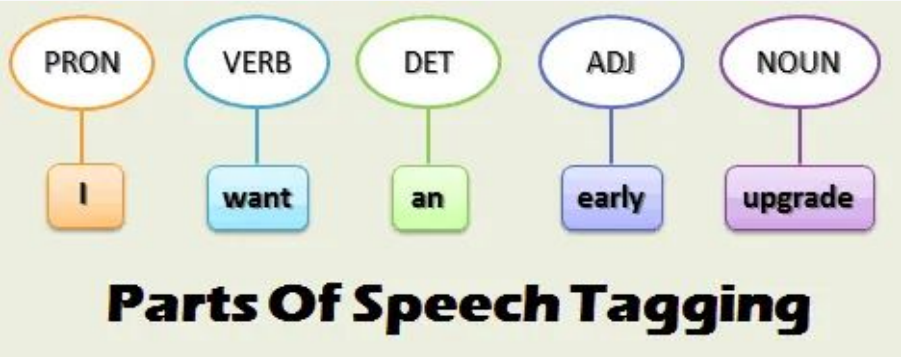
Common NLP tasks

- Tokenization
- **POS Tagging**
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- **POS** stands for **Part-of-Speech**, which is a linguistic term used to describe the **grammatical category of a word in a sentence.**
- **POS** tagging is the process of assigning each word in a text with its corresponding POS category, such as **noun, verb, adjective,** or **adverb**.
- **POS** tagging is a critical component in various natural language processing tasks, including **text-to-speech conversion, information retrieval,** and **machine translation**.

Common NLP tasks

- Tokenization
- **POS Tagging**
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



“John eats pizza.”

POS tagging would label "**John**" as a **proper noun** and "**eats**" as a **verb**, while syntactic parsing would identify "**John**" as the **subject** of the verb "eats" and "**pizza**" as the **object** of the verb.

POS tagging is concerned with the individual words, while syntactic parsing focuses on the overall sentence structure.

Common NLP tasks

- Tokenization
- **POS Tagging**
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Here are some more Part-of-Speech (POS) tagging examples:

Example 1: Simple Sentence

- Input: *"John plays football."*
- Output: John (NOUN) plays (VERB) football (NOUN).

Example 2: Sentence with Adjective and Adverb

- Input: *"The quick brown fox jumps over the lazy dog."*
- Output: The (DET) quick (ADJ) brown (ADJ) fox (NOUN) jumps (VERB) over (PREP) the (DET) lazy (ADJ) dog (NOUN)

Example 3: Ambiguous Word Example:

- Input: *"She will book a ticket."*
- Output: She (PRON) will (AUX) book (VERB) a (DET) ticket (NOUN).

Here, "book" is a verb, but in *"I read a book"*, "book" is a noun.

Common NLP tasks

- Tokenization
- POS Tagging
- **Word Sense Disambiguation**
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- **Word sense disambiguation** is the process of **determining the correct meaning of a word** when it has **multiple meanings**, based on context.
- This is a crucial task in natural language processing **because words often have different meanings depending on the context in which they are used**.
- Word sense disambiguation is used in various applications, including **information retrieval, machine translation, and question answering systems.**

Common NLP tasks

- Tokenization
- POS Tagging
- **Word Sense Disambiguation**
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

"The chicken is ready to eat."

Is the chicken cooked and ready for someone to eat, or is the chicken hungry and ready to eat?

Common NLP tasks

- Tokenization
- POS Tagging
- **Word Sense Disambiguation**
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Here are some examples of Word Sense Disambiguation (WSD):

Example 1 : Bank

- *He deposited money in the bank.* (Financial institution)
- *She sat by the bank of the river.* (Riverbank)

Example 2 : Bat

- *The bat flew out of the cave.* (Animal)
- *He hit the ball with a bat.* (Sports equipment)

Example 3 : Light

- *This box is very light to carry.* (Not heavy)
- *Turn on the light in the room.* (Illumination)

Example 4 : Apple

- *She ate a red apple.* (Fruit)
- *He works at Apple as a software engineer.* (Company)

Example 5 : Match

- *The football match was exciting.* (Game)
- *She lit a candle with a match.* (Fire-starting stick)

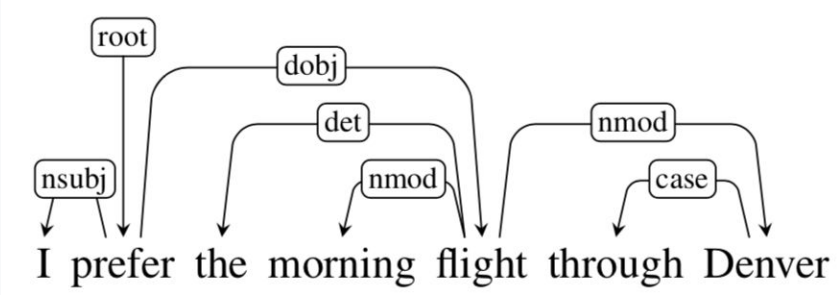
Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- **Dependency Parsing**
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Dependency parsing is the process of analyzing the **grammatical structure of a sentence** by identifying the **relationships between words in a sentence**.
- It involves identifying the **subject, object,** and **other dependent clauses** **and** **phrases**, and representing them as a **tree-like structure** known as a dependency tree.
- Dependency parsing is used in various natural language processing applications, including **sentiment analysis, named entity recognition, and machine translation**.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- **Dependency Parsing**
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



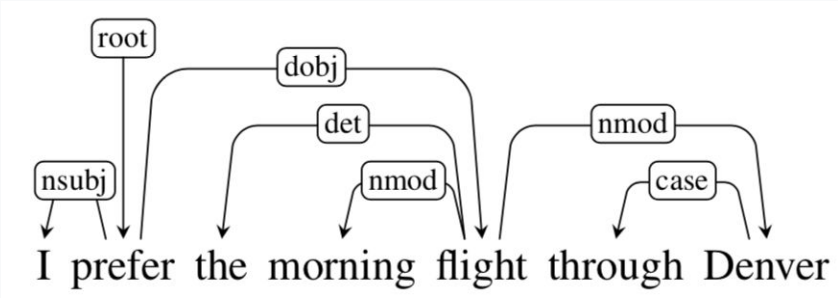
Head-Dependent: In the arrows representing relationship, the origin word is the Head & the destination word is Dependent.
For example: *(**I**, 'nsubj', **prefer**), 'prefer' is Head & 'I' is Dependent.*

Root: Word which is the root of our parse tree. It is 'prefer' in the above example.

Grammar Functions and Arcs: Tags between each Head-Dependent pair is a grammar function determining the relation between the Head & Dependent. The arrowhead carrying the tag is called an Arc.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- **Dependency Parsing**
- Syntactic Parsing
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	**Description**
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	**Description**
CONJ	Conjunct
CC	Coordinating conjunction

https://medium.com/data-science-in-your-pocket/dependency-parsing-associated-algorithms-in-nlp-96d65dd95d3e

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- **Syntactic Parsing**
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Syntactic parsing is the process of **analyzing the grammatical structure** of a sentence to determine its syntactic components, such as **nouns, verbs, adjectives,** and **adverbs.**
- It involves identifying the parts of speech of each word in the sentence and grouping them together into phrases and clauses based on their syntactic relationships.
- Syntactic parsing is used in various natural language processing applications, including **text-to-speech conversion, machine translation,** and **information retrieval.**

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- **Syntactic Parsing**
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

A sentence is structured as follows:

🔹 **S** = Noun Phrase (**NP**) + Verb Phrase (**VP**) + Prepositional Phrase (**PP**)

Example: ***The cat sat on the mat***

- '**The cat**' is the Noun Phrase (**NP**)
- '**sat**' is the Verb Phrase (**VP**)
- '**on the mat**' is the Prepositional Phrase (**PP**)

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- **Syntactic Parsing**
- Semantic Analysis
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

In English grammar, different groups of words exist based on specific rules. These include:

- **Noun Phrase (NP)** – It consists of a **Determiner (DET) + Nominal Noun (NN)**.
  For example, *'The big dog'* (*The* is the **DET**, and *big dog* is the **NN**

- **Verb Phrase (VP)** – It includes a **verb** and can have various combinations of objects and modifiers.

For example, *'is running quickly'* (***is running*** is the **VP** and the word **"quickly"** is an **adverb** because it modifies the verb **"running"** by describing *how* the action is performed

- **Prepositional Phrase (PP)** – This consists of a **Preposition (P) + Noun Phrase (NP)**.
  An example is *'on the table'*, where ***on*** is the **P** and ***the table*** is the **NP**.

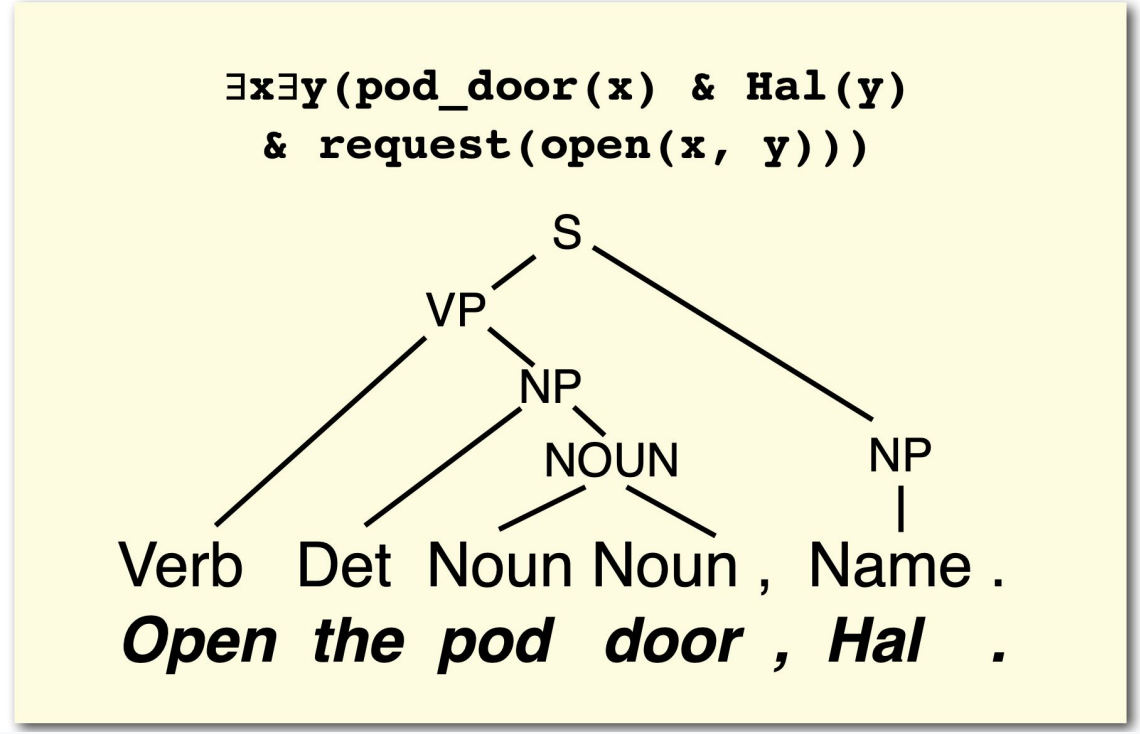
Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- **Semantic Analysis**
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Semantic analysis is the process of **extracting the meaning of a text** by analyzing the relationships between words and phrases in a sentence.
- It involves identifying the underlying concepts and ideas conveyed by the text and representing them in a structured form, such as a **knowledge graph or ontology**.
- Semantic analysis is used in various natural language processing applications, including **question answering, information retrieval,** and **chatbots**, to enable more accurate and intelligent responses.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- **Semantic Analysis**
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- **Semantic Analysis**
- Co-reference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

We need a meaning representation language.

"Shallow" semantic analysis: Template-filling (Information Extraction)

Named-Entity Extraction: Organizations, Locations, Dates,...
Event Extraction

"Deep" semantic analysis: (Variants of) formal logic

$\exists x \exists y(\text{pod\_door}(x) \& \text{Hal}(y) \& \text{request}(\text{open}(x,y)))$

We also distinguish between
Lexical semantics (the meaning of words) and
Compositional semantics (the meaning of sentences)

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- **Co-reference Resolution**
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Co-reference resolution is the task of **identifying all the expressions** (e.g., pronouns, names) in a text that refer to the same entity, and linking them together.
- It is a crucial task in natural language processing as it enables a system to maintain a **consistent representation of entities throughout a document**, enabling more accurate information extraction and text understanding.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- **Co-reference Resolution**
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Some examples of Co-reference Resolution are

Example 1: Sentence: *John bought a new car. He loves it.*
Coreference: "He" refers to "John," and "it" refers to "a new car."

Example 2: Sentence: *Sarah met Dr. Smith yesterday. The doctor gave her great advice.*
Coreference: "The doctor" refers to "Dr. Smith," and "her" refers to "Sarah."

Example 3: Sentence: *Elon Musk announced a new Tesla model. The billionaire shared details on Twitter.*
Coreference: "The billionaire" refers to "Elon Musk."

Example 4: Sentence: *My dog loves to run. Max always plays in the park.*
Coreference: "My dog" and "Max" refer to the same entity.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- **Coreference Resolution**
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

More than a decade ago, Carl Lewis stood on the threshold of what was to become the greatest athletics career in history. He had just broken two of the legendary Jesse Owens' college records, but never believed he would become a corporate icon, the focus of hundreds of millions of dollars in advertising. His sport was still nominally amateur. Eighteen Olympic and World Championship gold medals and 21 world records later, Lewis has become the richest man in the history of track and field -- a multi-millionaire.

Who is Carl Lewis?
Did Carl Lewis break any world records?
(and how do you know that?)

Common NLP tasks

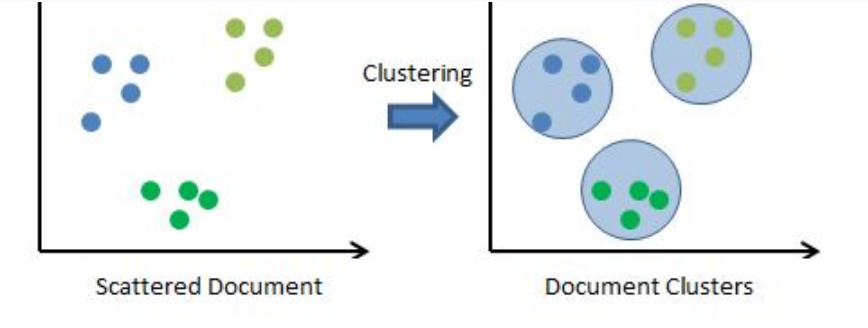
- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- **Named Entity Recognition (NER)**
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Named entity recognition (NER) is the process of identifying and categorizing named entities in a text, such as people, organizations, locations, and dates.

Figure 1: An example of NER application on an example text

Common NLP tasks

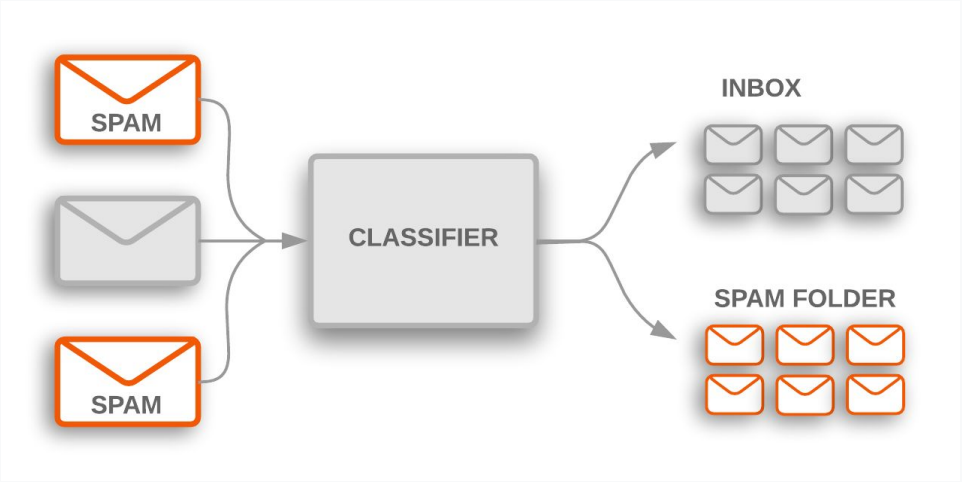
- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- **Text Representation**
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

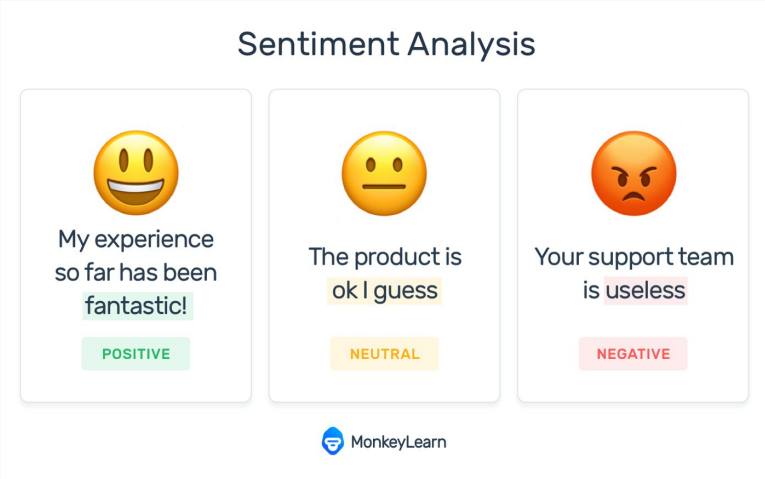


- Text representation is the process of converting **unstructured text data into a structured format** that can be used for natural language processing tasks.
- It involves selecting a suitable representation scheme, such as **bag-of-words, word embeddings**, or **topic models**, to capture the key features and characteristics of the text data in a **numerical form** that can be processed by machine learning algorithms.

Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- **Text Classification**
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



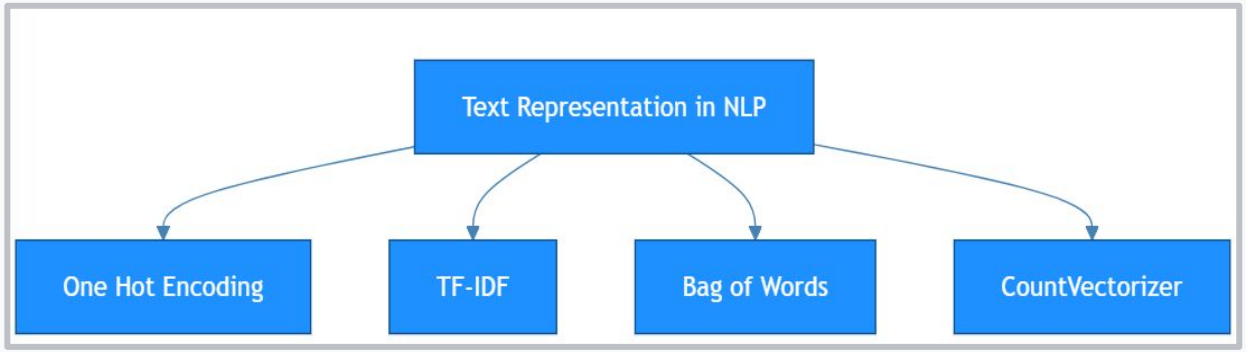


Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- **Natural Language Understanding**
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

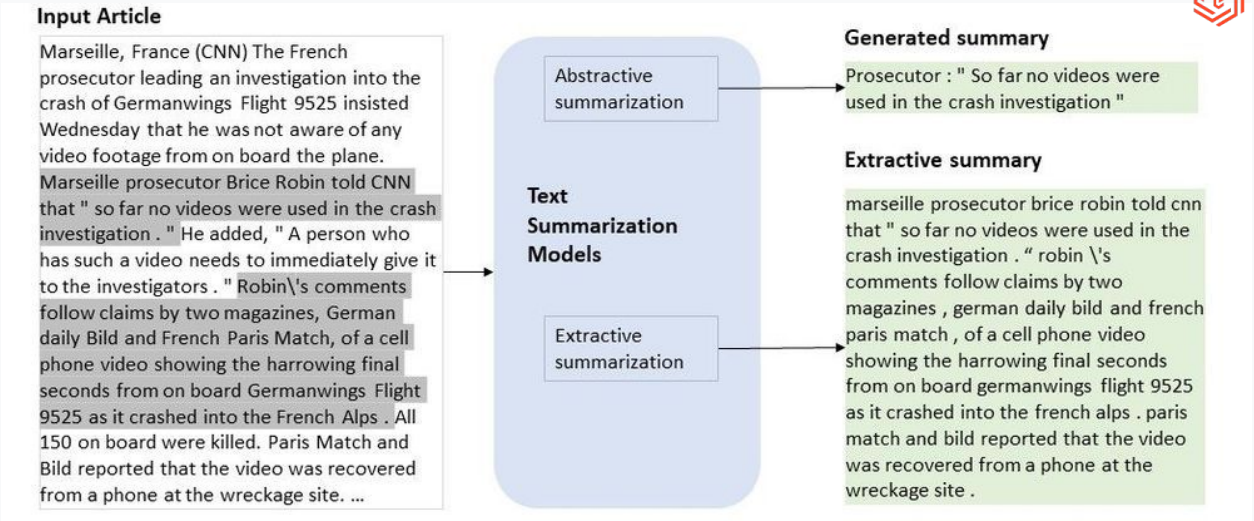
Natural language understanding

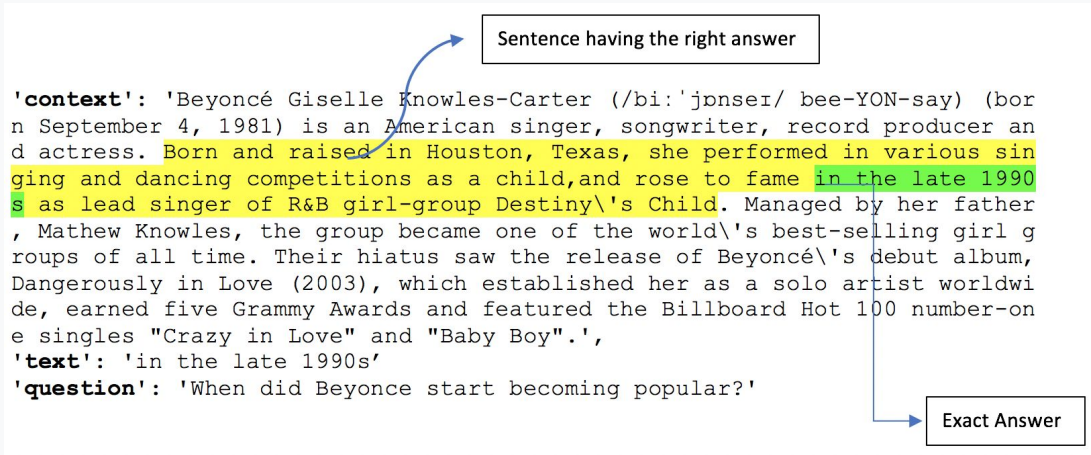
- Extract information (e.g. about entities or events) from text
- Translate raw text into a meaning representation
- Reason about information given in text
- Execute NL instructions



Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- **Natural Language Generation**
- Natural Language Translation
- Multimodal NLP





Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- **Natural Language Translation**
- Multimodal NLP

Natural language translation

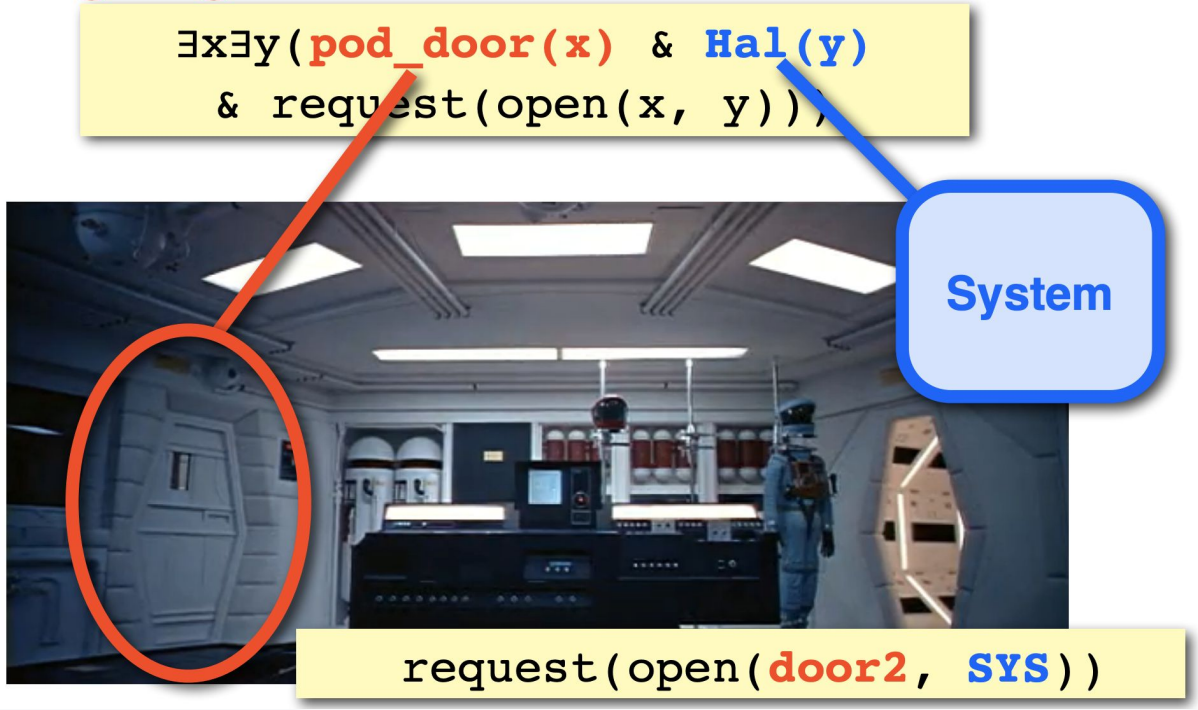
- Translate one natural language to another

https://www.shaip.com/blog/nlp-in-translation/

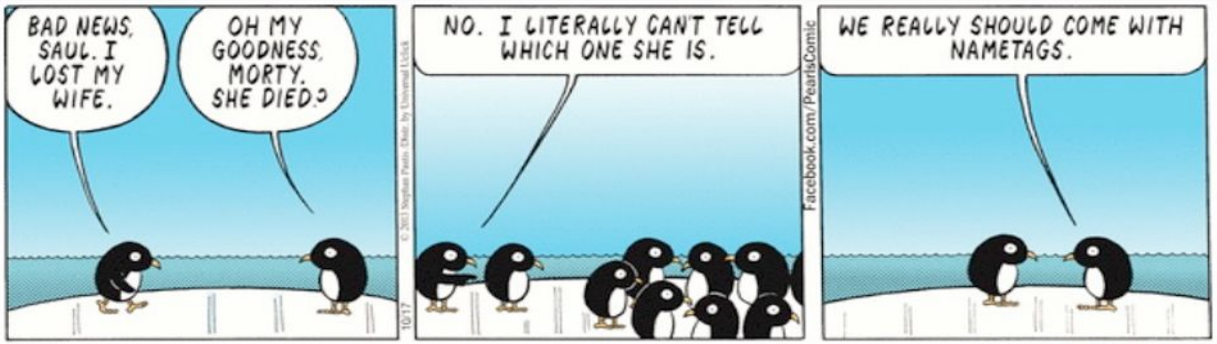
Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- **Multimodal NLP**

Multimodal NLP: mapping from language to the world



NLP Ambiguities



credit: A. Zwicky

NLP Ambiguities

1. Lexical Ambiguity: It is defined as the ambiguity associated with the meaning of a single word. A single word can have different meanings. Also, a single word can be a noun, adjective, or verb. For example, The word “bank” can have different meanings. It can be a financial bank or a riverbank. Similarly, the word “clean” can be a noun, adverb, adjective, or verb.

NLP Ambiguities

2. Syntactic Ambiguity: It is defined as the ambiguity associated with the way the words are parsed. For example, The sentence "Visiting relatives can be boring." This sentence can have two different meanings. One is that visiting a relative's house can be boring. The second is that visiting relatives at your place can be boring.

NLP Ambiguities

3. Semantic Ambiguity: This occurs when the meaning of words in a sentence can be interpreted in multiple ways. For instance, consider the sentence, *"Mary knows a little French."* Here, the phrase *"a little French"* is ambiguous because it could refer either to the French language or to a person of French origin.

Conclusion

What is NLP?

Key challenges and applications

Common NLP tasks

NLP Ambiguities